

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:30:39 ; Search time 4075.18 Seconds  
(without alignments)  
439.418 Million cell updates/sec

Title: US-09-300-482-14

Perfect score: 410

Sequence: 1 ccacgcgtccgcggtcatg.....gcaacggccgcgtcttgatg 410

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pli.\*

8: gb\_pl2.\*

9: gb\_pri.\*

10: gb\_pr2.\*

11: gb\_pr3.\*

12: em\_fun.\*

13: em\_hum1.\*

14: em\_hum2.\*

15: em\_in.\*

16: em\_om.\*

17: em\_or.\*

18: em\_ov.\*

19: em\_pat.\*

20: em\_ph.\*

21: em\_pl.\*

22: em\_ro.\*

23: em\_sts.\*

24: em\_sy.\*

25: em\_un.\*

26: em\_vi.\*

27: gb\_htg1.\*

28: gb\_htg2.\*

29: gb\_inl.\*

30: gb\_in2.\*

31: em\_bal.\*

32: em\_ba2.\*

33: em\_hum3.\*

34: em\_hum4.\*

35: gb\_pr4.\*

36: gb\_htg3.\*

37: gb\_htg4.\*

38: gb\_htg5.\*

39: gb\_htg6.\*

40: gb\_htg7.\*

41: em\_htg1.\*

42: em\_htg2.\*

43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_vil.\*  
94: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	394.8	96.3	1686	7	AF061838
2	389.4	95.0	634	7	AF037030
3	326	79.5	150120	8	AF001552
4	323.8	79.0	2392	7	AF037037
5	320.6	78.2	1818	7	AF061837
6	319	77.8	645	7	AF037029
7	262	63.9	96240	56	AC068900
8	230	56.1	1667	7	AB007907
9	216.2	52.7	1618	45	MSU18239
10	195.2	47.6	3158	8	LDI130772
11	183	44.6	111222	7	AC007764
12	175	42.7	79837	7	AB005233

c

Sat Nov 4 18:11:08 2000

13	114.8	28.0	38587	29	AC011913	AC011913 Leishmani
14	110.4	26.9	42000	57	AC073566	AC073566 Leishmani
15	109.4	26.7	1683	8	D89161	D89161 Schizosacch
16	106.8	26.0	2156	54	TBGNNG	TBGNNG T. brucei gn
17	99.2	24.2	1916	5	E13660	E13660 gDNA encodi
18	96	23.4	23773	1	AE002301	AE002301 Chlamydia
19	94.4	23.0	1953	3	OAGPDGH	OAGPDGH Ovine 6-pgd
20	93	22.7	780	8	CNS01AT6	CNS01AT6 Botrytis
21	92.8	22.6	2853	45	SC6PGRDHY	SC6PGRDHY S. cerevisia
22	92.8	22.6	2853	45	SCU17155	SCU17155 Saccharomyc
23	92.8	22.6	10801	45	YSC9186	YSC9186 Saccharomyc
24	92.8	22.6	16854	1	AE001722	AE001722 Thermotog
25	91.2	22.2	14168	1	AE001281	AE001281 Chlamydia
26	89.8	21.9	10195	65	U32737	U32737 Haemophilus
27	88	21.5	9940	1	AE002201	AE002201 Chlamydia
28	88	21.5	10447	1	AE001620	AE001620 Chlamydia
29	86.6	21.1	39228	2	MLCB1788	MLCB1788 Mycobacte
30	85.8	20.9	36021	2	MTCY359	MTCY359 Mycobacte
31	82.4	20.1	2053	65	SYOPGD	SYOPGD Mycobacteri
32	82.4	20.1	2263	2	SSGNDG	SSGNDG Synecococc
33	82.2	20.0	1536	82	HSU30255	HSU30255 Human phosp
34	82.2	20.0	6278	1	BACNT8L	BACNT8L Bacillus li
35	81.6	19.9	1610	7	AB006102	AB006102 Candida a
36	81.6	19.9	43325	45	SPBC660	SPBC660 S. pombe c
37	80.8	19.7	1335	2	EVU14465	EVU14465 Escherichia
38	80.4	19.6	13799	1	AE002359	AE002359 Treponema
39	80.2	19.6	14838	1	AE001213	AE001213 Actinobacil
40	78.4	19.1	14838	2	D88189	D88189 Klebsiella
41	77.4	18.9	10903	2	KPU14463	KPU14463 Klebsiella
42	77.2	18.8	1335	2	KPU14463	KPU14463 Klebsiella
43	76.8	18.7	125469	65	SYCSLRA	SYCSLRA Vibrio ch
44	75.2	18.3	10548	1	AE004417	AE004417 Vibrio ch
45	74	18.0	1335	2	CAU14426	CAU14426 Citrobacter

ALIGNMENTS

RESULT 1

LOCUS AF061838 1686 bp mRNA 08-JUN-2000

DEFINITION Zea mays strain W64A x A182E putative cytosolic 6-phosphogluconate dehydrogenase (pdh2) mRNA, complete cds.

ACCESSION AF061838

VERSION AF061838.1

KEYWORDS GI:3342801

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 1686)

AUTHORS Redinbaugh,M.G. and Campbell,W.H.

TITLE Nitrate regulation of the oxidative pentose phosphate pathway in maize (Zea mays L.) root plastids: induction of 6-phosphogluconate dehydrogenase activity, protein and transcript levels

JOURNAL Plant Sci. 134 (2), 129-140 (1998)

REFERENCE 2 (bases 1 to 1686)

AUTHORS Redinbaugh,M.G. and Bailey-Serres,J.

TITLE Characterization of cDNAs encoding cytosolic and a putative plastidic 6-phosphogluconate dehydrogenase

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1686)

AUTHORS Redinbaugh,M.G., Campbell,W.H. and Bailey-Serres,J.

TITLE Direct Submision

JOURNAL Submitted (28-APR-1998) USDA, ARS/Dept.. Plant Path., OARDC/OSU, 1680 Madison Ave., Wooster, OH 44691, USA

FEATURES

source

gene

99. .1547

/gene="pdh2"

/note="NADPH producing dehydrogenase of the oxidative pentose phosphate pathway; PGD"

/codon\_start=1

/product="putative cytosolic 6-phosphogluconate dehydrogenase"

/protein\_id="AAC27703.1"

/db\_xref="GI:3342802"

/translation="MALTRIGLAGLVGMQNALNIAEKGPPIISVYNRRTTSKVDETQY RAKAENGLPVYGFHDPASFKSIQPRVIMLVKAGAPVDTIATLAHLEQGCIIID RAGNENRTERKAMEERGLLYLGMVVGEEGARGNPSLMPGSGSEAFKYKVDIVL KVAAGPDSQPCVYIGKSGSNFVKVHNGIEYDGMQIIEADYDKLFKFGKLTNSEL HGVSEWNKGSLLEFLITADIFGIDKHDGDIYDKMLDKTGKGTGQVAAAE LSVAAPTEASDLSFLSLGDLDERVEAKIFQGVGYSPVDAQVLEVDNRNPGLAS KICSTAQGMNIIKAKSAEKGKGLNIGELARIWKGCIIIRAIFLDRIKKAIDVRNPGLAS LIVDPEFAQEIIMDRQANRRVVCILAINGVSTPGMSASIAFYDSYRDRRLPANLVQAO RDIYGAHTYERYDMPGFSFHTFKTARNISN"

BASE COUNT 367 a 462 c 535 g 322 t

ORIGIN

Query Match 96.3%; Score 394.8; DB 7; Length 1686;

Best Local Similarity 99.2%; Pred. No. 1.7e-71;

Matches 396; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 12 qcgtcatggggcagaaaccttgcctcaacattgcagagaaaggggtccccatctctgtg 71

Db 132 GCGGTGATGGGGCAGACCTTGCCCTCAACATTGCAGAGAAAGGGTTCCCATCTCTGTG 191

QY 72 tacaacaggacaacctccaaggtggagagaccgtgcagcgtcccaagcgacagagaac 131

Db 192 TACAACAGGACAACCTCCAAGGTGGAGAGACCGTGCAGCGTGCCCAAGGAGAGGAAC 251

QY 132 ottccgtctacgcttccatgaccccgctcctttgtgaagtccattcagaagccacgg 191

Db 252 CTTCCGCTACGGCTTCCATGACCCCGCGCTCTTTGTGAAGTCCATTCAGAAAGCCACGG 311

QY 192 gtggtgatcatgctgtcaaggcggcgccagttgacacagaccatcgagcgtcgca 251

Db 312 GTGGTGATGATGCTCGTCAAGCGCGGCGCCAGTTGACACAGACCATCGGACGCTCGCA 371

QY 252 gctacttggagcaggcgactgcatactgcattgagggggaacagtggtgcagaaacag 311

Db 372 GCTCACTTGGAGCAGGCGGACTGTCATCATCGATGGGGGGAACAGTGGTACGAAACACG 431

QY 312 gagagaggagaaagccatgagagcgcgccctnctgtatcttgcgcatgggtgtctct 371

Db 432 GAGAGAGGAGAGAGCCATGATGAGGAGCGCGGCTCTCTGTATCTGGCATGTTGTCTCT 491

QY 372 ggaggaagaggagggtgcccgaacggcccgctccttgatg 410

Db 492 GGAGGAGAGAGAGGGTGCCCGCAACGGCCGCTCTTGATG 530

RESULT 2

AF037030 634 bp mRNA PLN 26-NOV-1998

LOCUS Zea mays 6-phosphogluconate dehydrogenase isoenzyme-B mRNA, partial cds.

DEFINITION

ACCESSION AF037030

VERSION AF037030.1

KEYWORDS GI:3925224

SOURCE Zea mays.

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 634)

AUTHORS Padegimas,L.S. and Reichert,N.A.

TITLE Isolation of genes and regulatory sequences implicated in hypersensitive response from Zea mays nematode-resistant line MP3070

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 634)

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3925224

26-NOV-1998

PLN

26-NOV-1998

AF037030

AF037030.1

GI:3

## ALIGNMENTS

RESULT	1	AF061838	1686 bp	mrna	PLN	08-JUN-2000
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						

# AUTHORS TITLE JOURNAL

Padegimas, L.S. and Reichert, N.A.  
Direct Submission  
Submitted (06-DEC-1997) Department of Plant and Soil Sciences,  
Mississippi State University, Box 9555, Mississippi State, MS  
39762, USA

## FEATURES source

Location/Qualifiers  
1. 634  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
100. .5634  
/note="6pghd(B)"  
/codon\_start=1  
/product="6-phosphogluconate dehydrogenase isoenzyme B"  
/protein\_id="AAC79950.1"  
/db\_xref="GI:392525"  
/translation="MALTRIGLAGLVGONLALNTAEKGFPIVSYNRTTSKYDETQV  
RAKAGNLPYIGFHPASFNOSIKPRVIMLVKAGAPVDQTATLAHLEQDCDIID  
GNGRYENTERREKAMEBGLLYLGMVSGGEGGARGPSLMPGGSFDAKYVEDIVL  
KVAQVPDPSGCVTYIGK"  
BASE COUNT 139 a 174 c 204 g 117 t  
ORIGIN

Query Match 95.08; Score 389.4; DB 7; Length 634;  
Best Local Similarity 98.2; Pred. No. 2.8e-70;  
Matches 393; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 11 cgcggatcagggcagaaacctcgaaggtggagcagacgttcagagaggggtcccatctctgt 70  
Db 132 CGCTGTCTATGCGGAGAACCTTCCCTCAACATTGCAGAGAAAGGGTCCCATCTCTGT 191  
Qy 71 gtacaacaggagaaacctcgaaggtggagcagacgttcagagaggggtcccatctctgt 130  
Db 192 GTACAACAGGAGAACCTTCCCTCAACATTGCAGAGAAAGGGTCCCATCTCTGT 251  
Qy 131 cctcccgctacggttcagatcagcccgcttcctttgtgaagtcattcagaagccagc 190  
Db 252 CCCTCCCGCTACGCTTCCATGACCCCGCTCTTGTGAAGTCCATTCAGAACCCACG 311  
Qy 191 ggtggtgatcgtctcagagcagcgcgcagctgacagacacacacacacacacacac 250  
Db 312 GGTGATGATCATGCTCTGTCAGGCGCGCGCCAGTTTGACACACACATCGCGACGCTCG 371  
Qy 251 agtcaacttgagcagggcgactgcacatcagatgggggagcagctggtacagagacac 310  
Db 372 AGCTCACTTGAGCAGGCGGACTGCATCATCGATGCGGAGGAGCGGTACGAGAACAC 431  
Qy 311 ggagagagagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 370  
Db 432 GGAG 491  
Qy 371 tgaggagagagagagagagcagcagcagcagcagcagcagcagcagcagcagcagcag 410  
Db 492 TGGAG 531

## RESULT 3 AP001552

LOCUS AP001552 150120 bp DNA PLN 16-JUN-2000  
DEFINITION Oryza sativa genomic DNA, chromosome 6, PAC clone: P0029D06.  
ACCESSION AP001552  
VERSION AP001552.1 GI:7363267

## KEYWORDS SOURCE ORGANISM

Oryza sativa (cultivar: Nipponbare) DNA, clone: P0029D06.  
Oryza sativa  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

## REFERENCE AUTHORS TITLE

1 (bases 1 to 150120)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare (GA3), genomic DNA, chromosome 6, PAC  
clone: P0029D06  
Published Only in Database (2000) In press  
REFERENCE 2 (bases 1 to 150120)

## JOURNAL REFERENCE

Join(13886. .13953.14703. .14914.14920. .15114.15144. .15314.15374. .15492)

# AUTHORS TITLE JOURNAL

Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (29-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji  
Sasaki, National Institute of Agrobiological Resources, Rice Genome  
Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@agr.affrc.go.jp,  
URL: http://www.dna.affrc.go.jp:82/, Tel: 81-298-38-7441,  
Fax: 81-298-38-7468)

## COMMENT

The orientation of the sequence is from SP6 to T7 of the PAC clone.  
Genes were predicted from the integrated results of the  
following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as  
SplicePredictor (October 1998 version). The genomic sequence was  
searched against the non-redundant database NRP (PIR, SWISSPROT,  
GENPEPT, PDB) from MAF DNA bank and the cDNA sequence database at  
RGP. Protein similarities of the coding regions were searched  
against NRP with BLASTP2.0. ESTs represent the identified cDNA  
sequences using BLASTN2.0 with the corresponding DDBJ accession no.  
and RGP clone ID.  
This sequence of this clone has an overlap with P0541H01 clone,  
DDBJ: AP001389 at the 3' end.  
This clone ends at the position 26,826 of P0541H01. Detailed  
information on overlap and assembly quality together with  
annotation of this entry at  
http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.

## FEATURES source

Location/Qualifiers  
1. 150120  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="6"  
/clone="P0029D06"  
complement(join(1887. .2096.2316. .2390))  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA93012.1"  
/db\_xref="GI:7363268"  
/translation="MIKAPIAVTVLGRSLCEGLAAGKRGDCCEANRVSGICPVADVE  
RKVVEVAGIEGSRWCRGLGAEVLSWRRAGLAWRPAAGGRLARHEA"  
join(2761. .3848.5363. .5544.5878. .5959.6846. .7125)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA93013.1"  
/db\_xref="GI:7363269"  
/translation="MARRKGLDLSQPDILLHILSNVRYKEAVRTAAVSRWKLHL  
TKLPALSFIMSVLGAQSSLSQSRQVDSMARTLRRCAGPDRTVKRLCLAVRKDV  
PMCRYADEFIATAAASGLFLNCPKLNDDAPWSLHLPAATACLSWESYVSR  
PPVHVGPCASALKSLTFKDFMVLHPCYLODTAPPSLEELHISCTLSGSIETSATM  
PRKLRIADYVSVSLCTAAIAVLADIELTLRVSHDGGKPPSPSHMLCVETLFR  
ASFTESYFRLAPKLVFDMRCCKYKREVRDVAVRHLSDVYIELFAGRLPCYNKAK  
RFLQMEDCDLMDILQGIIMPGRKKYVOSLPYRRANKVRWEGQDDREVEEVQWHPR  
MDMLTTMAASAGPGDCTAASDGVYMAATARAGEDGNSDLGSGVMEDLIGAAARW  
RPADARRRMFGQRRRGNEAVTGWARRAERYSVTAGGSTSARLRRPDAAR  
TSGSAHEDEPDGEGGAGAEEGAGSFSYKFKON"  
join(8143. .8539.8557. .8908.9092. .9284)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA93014.1"  
/db\_xref="GI:7363270"  
/translation="MVVDRKPAVPLVLRPPASRCRLDRRYAEQLLPKRGKAPAG  
EGEGVRRASAGGAGSHIRCPSSPTLLGKSGTSPGVRGREGSGIGEMKSM  
APTSVELLRASAGAEVELAGSGGSLTGKSEVVAHEEGARHRIQAAVSSAA  
TPGPDGHRVPPADVTTRKIRREAPASGPTDVVCAAMRRLGLVAAVAPPPPPR  
TPHRRHLPTSTFNHRRHYPCWLRRSREKRAAEVEEQRAAETPLPASEVRENK  
SSDARHAGRLRRRGQGRVPRAGEHRTMAVTVEGH"  
join(11170. .11233.13292. .13386)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA93015.1"  
/db\_xref="GI:7363271"  
/translation="MNLKVHPKGTGKVGANGSDKGLDGSDETNLFTSSTKDFNS  
GAGKREQD"  
join(13886. .13953.14703. .14914.14920. .15114.15144. .15314.15374. .15492)





## RESULT 4

AF037037 2392 bp DNA PLN 26-NOV-1998  
 LOCUS Zea mays 6-phosphogluconate dehydrogenase isoenzyme A gene, partial  
 DEFINITION cds.  
 ACCESSION AF037037  
 VERSION AF037037.1 GI:3925238  
 KEYWORDS Zea mays.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 REFERENCE 1 (bases 1 to 2392)  
 AUTHORS Padegimas, L.S. and Reichert, N.A.  
 TITLE Isolation of genes and regulatory sequences implicated in  
 hypersensitive response from Zea mays nematode-resistant line MP307  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2392)  
 AUTHORS Padegimas, L.S. and Reichert, N.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-DEC-1997) Department of Plant and Soil Sciences,  
 Mississippi State University, Box 9555, Mississippi State, MS  
 39762, USA  
 FEATURES Location/Qualifiers  
 source 1..2392  
 /organism="Zea mays"  
 /db\_xref="taxon:4577"  
 1858..>2392  
 /note="6pgh(A)"  
 /codon\_start=1  
 /product="6-phosphogluconate dehydrogenase isoenzyme A"  
 /protein\_id="AAC79957.1"  
 /db\_xref="GI:3925239"  
 /translation="MALTRIGLAGLVGMGNLALNIAEKGPPISVYNTTTSKVDETVQ  
 RAKVEGNLPVFGFHPDPASFVSSIOKPRVIMLVKAGAPVDOTIATLAHLDQGDICVD  
 GGNWENTREKAMEERGLLYLGMVSGGEGARNGPSLMPGGSEFAYKYIEDILL  
 KVAQVPDPSGFCVYIGKSGSGNFVKMVHNGIEYGMQIAEAYDVLKSGYKLTNSEL  
 HOVSEWNGKGLLSFLIEITADIFGDKDGEGLYVDKLDKTMGTGKWTQVQAAE  
 LSVAAPTLEASLDLSRFLSGLDKDERVEASKIPQGDYSTGLPDKRAQLIEDYRQALYASK  
 ICSYAQGNIIKAKSSSEKGLNGLNGLAIWKGCCIIRALFLDKIKAYDRNPLANL  
 LVDPFAQEIIDRQAAWRRVVLAINNGVSTPGMSASLAYFDSYRRDLPLANLVQQR  
 DYFGAHTVERVDMPLVLTPELRETTSALTREKV"

## BASE COUNT

642 a 529 c 546 g 668 t 7 others

## Query Match

Best Local Similarity 79.0%; Score 323.8; DB 7; Length 2392;

Matches 352; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 11 cgcggtcatgggagaaacattgcctcaacattgcagagaagggttccccatctgt 70  
 DB 1890 CGCGGTATGGGACAGAACCTTGCCTCAACATCGGAGAAAGGTTCCCATCTCGT 1949  
 QY 71 gtacaacagagaacacctcaaggtagacagaccgtgcagcgtgccaaaggcagaagaaa 130  
 DB 1950 CTACAACAGAGACACCTCCAAAGTTGATGAGACCGTGCAGCGTGCAGAGGAA 2009  
 QY 131 ccttcccgctacggtcttcattgaccccgctcttcttggaagtccattcagaagccag 190  
 DB 2010 CTTCCAGTGTGTTGGTTTCCACAGCCCCCGCTCTCTGTGATGCTCCATCAGAGCCCG 2069  
 QY 191 ggtggtatctctgcgtcaaggccggcgccagtgtagccagaccatcgagcgtcgc 250  
 DB 2070 TGTGCTGATCATGCTGCTCAAGGCTGGGGCGCGGTGACGACGACATTCGACGCTCG 2129  
 QY 251 agctcacttggagcagcgactgcattcagatggggggaacagagtggtcagaacac 310  
 DB 2130 GCGCACCTTGATCAGGGGACTGATGCTGATGTTGGCAACAGAGTGGTACGAGAACAC 2189  
 QY 311 ggaagaggaggaaggccattgagagcgcgccctnctgtatcttgcatgggtgtctc 370  
 DB 2190 GGAGAGGAGGAGAGGCGGATGAGAGCGGGGCTCTATATCTTGCAATGGCGCTC 2249  
 QY 371 tgagagaaaggaggtgccgcaacggccgctcttgatg 410  
 DB 2250 CGGAGAGAGGAGGGTGGCCCGCAATGGCCGCTCTTGATG 2289

## RESULT 5

AF061837 1818 bp mRNA PLN 08-JUN-2000  
 LOCUS Zea mays strain W64A x A182E putative cytosolic 6-phosphogluconate  
 DEFINITION dehydrogenase (pdhl) mRNA, complete cds.  
 ACCESSION AF061837  
 VERSION AF061837.1 GI:3342799  
 KEYWORDS Zea mays.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 REFERENCE 1 (bases 1 to 1818)  
 AUTHORS Redinbaugh, M.G. and Campbell, W.H.  
 TITLE Nitrate regulation of the oxidative pentose phosphate pathway in  
 maize (Zea mays L.) root plastids: induction of 6-phosphogluconate  
 dehydrogenase activity, protein and transcript levels  
 JOURNAL Plant Sci. 134 (2), 129-140 (1998)  
 REFERENCE 2 (bases 1 to 1818)  
 AUTHORS Redinbaugh, M.G. and Bailey-Serres, J.  
 TITLE Characterization of cDNAs encoding cytosolic and a putative  
 plastidic 6-phosphogluconate dehydrogenase  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1818)  
 AUTHORS Redinbaugh, M.G., Campbell, W.H. and Bailey-Serres, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-APR-1998) USDA, ARS/Dept. Plant Path., OARDC/OSU,  
 1680 Madison Ave., Wooster, OH 44691, USA  
 FEATURES Location/Qualifiers  
 source 1..1818  
 /organism="Zea mays"  
 /strain="W64A x A182E"  
 /db\_xref="taxon:4577"  
 1..1818  
 /gene="pdhl"  
 69..1523  
 /gene="pdhl"  
 /note="NADPH producing dehydrogenase of the oxidative  
 pentose phosphate pathway."  
 /codon\_start=1  
 /product="putative cytosolic 6-phosphogluconate  
 dehydrogenase"  
 /protein\_id="AAC27702.1"  
 /db\_xref="GI:3342800"  
 /translation="MALTRIGLAGLVGMGNLALNIAEKGPPISVYNTTTSKVDETVQ  
 RAKVEGNLPVFGFHPDPASFVSSIOKPRVIMLVKAGAPVDOTIATLAHLDQGDICVD  
 GGNWENTREKAMEERGLLYLGMVSGGEGARNGPSLMPGGSEFAYKYIEDILL  
 KVAQVPDPSGFCVYIGKSGSGNFVKMVHNGIEYGMQIAEAYDVLKSGYKLTNSEL  
 HOVSEWNGKGLLSFLIEITADIFGDKDGEGLYVDKLDKTMGTGKWTQVQAAE  
 LSVAAPTLEASLDLSRFLSGLDKDERVEASKIPQGDYSTGLPDKRAQLIEDYRQALYASK  
 ICSYAQGNIIKAKSSSEKGLNGLNGLAIWKGCCIIRALFLDKIKAYDRNPLANL  
 LVDPFAQEIIDRQAAWRRVVLAINNGVSTPGMSASLAYFDSYRRDLPLANLVQQR  
 DYFGAHTVERVDMPLVLTPELRETTSALTREKV"

## BASE COUNT

418 a 474 c 520 g 406 t

## Query Match

Best Local Similarity 78.2%; Score 320.6; DB 7; Length 1818;

Matches 350; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 11 cgcggtcatgggagaaacattgcctcaacattgcagagaagggttccccatctgt 70  
 DB 101 CGCGGTATGGGACAGAACCTTGCCTCAACATCGGAGAAAGGTTCCCATCTCGT 160  
 QY 71 gtacaacaggaacacctcaaggtagacagaccgtgcagcgtgccaaaggcagaagaaa 130  
 DB 161 CTACAACAGGACGACCTCCAGGTGATGAGACCGTGCAGCGTGCAGAGGTCGAAGAAA 220  
 QY 131 ccttcccgctacggtcttcattgaccccgctcttcttggaagtccattcagaagccag 190  
 DB 221 CTTCCCGCTGTTGGTTTCCACGACCCCGCTCTCTGCTGAGCTCCATCCAGAGCCCCG 280



\* 50875 56713: contig of 5839 bp in length  
 \* 56714 56763: gap of unknown length  
 \* 56764 61235: contig of 4472 bp in length  
 \* 61236 61285: gap of unknown length  
 \* 61286 66859: contig of 5574 bp in length  
 \* 66909: gap of unknown length  
 \* 66910 73521: contig of 6612 bp in length  
 \* 73522 73571: gap of unknown length  
 \* 73572 78793: contig of 5222 bp in length  
 \* 78794 78843: gap of unknown length  
 \* 78844 83051: contig of 4208 bp in length  
 \* 83052 83101: gap of unknown length  
 \* 83102 96240: contig of 13139 bp in length.

## FEATURES

Location/Qualifiers  
 1..96240  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /chromosome="III"  
 /clone="IGF-F1A12"  
 BASE COUNT 28496 a 18610 c 20275 g 28013 t 846 others  
 ORIGIN

Query Match 63.9%; Score 262; DB 56; Length 96240;  
 Best Local Similarity 78.4%; Pred. No. 1e-44;  
 Matches 313; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 12 gcggtcatggggcagaaccttgcctcaacattcagagaagggttcccatctctgtg 71  
 Db 13068 GCTGTGATGGGTCAGAACTAGCTCTCAACATTGCAGAAAGGCTTCCCAATCTCTGT 13127  
 QY 72 tacaacaggacaacctccaagtgagagaccctgcagctgcgaaggcagagaagaac 131  
 Db 13128 TACACAGAACTACTTCAAAAGTTGATGAGACCTCGAACGAGCCAAAGAGGCAAT 13187  
 QY 132 ctcccgctacgcttcacatgaccccgctccttcttgagtcattccatcagaagcag 191  
 Db 13188 CTTCTCTTTACGGCTTCATGACCCCGAGTCTTTTCTCAAGTCCATTCAGAGCCACGC 13247  
 QY 192 gtggtgatcatgctcgtcaaggcgcgcgcagctgaccagacatcgagcagcgtcga 251  
 Db 13248 GTTATTATCATGCTGTGTAAGCTGGTTCCTCCGCTGACAGACCATCAAGACCTCTCT 13307  
 QY 252 gctcacttgagcaggcgcagctcatcgatgggggaacagagtggtacgagaacacg 311  
 Db 13308 GCTTATTGGAGAAAGTGATGATGCTGATGATGAATGATGATGATGATGATGATGAT 13367  
 QY 312 gagagaggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 371  
 Db 13368 GAGAGGAG 13427  
 QY 372 ggaggaaggagggtgcccgaacgcccgcctctcttgatg 410  
 Db 13428 GGTGGTGAAGAAGGTGCTGCTAATGGCCCATCATGATG 13466

## RESULT 8

AB007907  
 LOCUS  
 DEFINITION  
 Glycine max mRNA for 6-phosphogluconate dehydrogenase, complete cds.

AB007907  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AB007907.1 GI:2529228  
 6-phosphogluconate dehydrogenase; gnd.  
 Glycine max root nodules cDNA to mRNA.  
 Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 Rosidae; Fabales; Fabaceae; Papilionoideae; Glycine.  
 1 (bases 1 to 1667)  
 Katsurada, A.  
 Direct Submission  
 JOURNAL Submitted (07-OCT-1997) to the DDBJ/EMBL/GenBank databases. Akihiko

Katsurada, Tezukayama Gakuin College, Dept. of Nutrition; 4-2-2  
 Harumidai, Sakai, Osaka 590-01, Japan  
 (E-mail: LDD00547@niftyserve.or.jp, Tel: +81-722-96-1331,  
 Fax: +81-722-92-2135)  
 2 (bases 1 to 1667)  
 Katsurada, A.

## REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source

Location/Qualifiers  
 1..1667  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /tissue\_type="root nodules"  
 55..1590  
 /gene="gnd"  
 55..1590  
 /gene="gnd"  
 /EC\_number="1.1.1.44"  
 /codon\_start=1  
 /product="6-phosphogluconate dehydrogenase"  
 /protein\_id="BAA22812.1"  
 /db\_xref="GI:2529229"  
 /translation="MAQFSTGIGLAGLAVMGONLALNAEKGFFISVYNRRTSKVDET  
 VGRKQEGNLPVGYGHDPEAFVHSIQPRVIMLVKAGAPVDQITKISAYMEKGDGI  
 IDGNEWENTERREREKVAELGLYLGLVSGGEEGARGNSLMPGGSFYFAFKYLEDI  
 LLKVAQVPDSCPVYIGKGGSNFVKMHNGIEYGMOLIAEAYDLKSVGKLSNE  
 ELQSVFSEWNKCELLSFLIEITADIFGDKDKDGVLVDKLDTKMGKTGKWTVOQA  
 AELSIAAPTIEASLDARLSGLKEERVEAAKVFSGGIGDIVDHDVQKLDIDVRK  
 ALYAKICSYAOGMNLIRAKSTEKGWDLKELARLWKGGCIIIRAFIDRLIKQAYERN  
 PNLALLVDPEFAKIIDYQSARVRVCLAINSGISTFGMSASLAYFDTRYRRLPAN  
 LVQAQRDYFGAHTYERVDIEGSHTEWFKLAKQSKLLDYCISANQDFPNKCNIFCSDC  
 MLS"  
 1667  
 /note="34 A and 1 G nucleotides"  
 BASE COUNT 480 a 277 c 415 g 495 t  
 ORIGIN

Query Match 56.1%; Score 230; DB 7; Length 1667;  
 Best Local Similarity 73.4%; Pred. No. 9.2e-38;  
 Matches 293; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
 QY 12 gcggtcatggggcagaacacctccctcaacattgcagagaagggttcccatctctgtg 71  
 Db 94 GCTGTATGGGCCAAATCTAGCACTCAATATGCTGAGAAAGGCTTCCCATCTCTGT 153  
 QY 72 tacaacaggacaacctccaagtgagcagaccctgcagctgcgaaggcagagaagaac 131  
 Db 154 TATAACCGAACCACCTCCAAAGTTGATGAGACTGTAGAACGAGCAAAACAAAGGAAAT 213  
 QY 132 ctcccgctacgcttcacatgaccccgctccttcttgagtcattccatcagaagcagcag 191  
 Db 214 CTTCCAGTTTATGGCTACCATGACCCCGAAGCTTTTGTTCATTCCATCAAAAGCCTAGG 273  
 QY 192 gtggtgatcatgctcgtcaaggcgcgcgcagctgaccagacatcgagcagcagcagcag 251  
 Db 274 GTGATAATAATGCTGTGTTAAGCTGGGCACCTGTTGACAGACCATTAAGACCTATCT 333  
 QY 252 gctcacttgagcaggcgcagctgcacatcgatgggggaacagagtggtacgagaacacg 311  
 Db 334 GCATACATGGAAGAAAGGTGACTGTATAATTATGATGGTGTGAACGAATGATGAGAACT 393  
 QY 312 gagagaggagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 371  
 Db 394 GAAAG 453  
 QY 372 ggaggaaggagggtgcccgaacgcccgcctctcttgatg 410  
 Db 454 GGTGGTGAAGAAGGTGCTGTAATGGTCCCTCTTTGATG 492

RESULT 9

MSU18239 1618 bp mRNA PLN 30-JAN-1997  
LOCUS Medicago sativa 6-phosphogluconate dehydrogenase mRNA, complete  
DEFINITION cds  
ACCESSION U18239  
VERSION U18239.1 GI:603220  
KEYWORDS  
SOURCE alfalfa.  
ORGANISM Medicago sativa subsp. sativa  
Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
Fabales; Fabaceae; Papilionoideae; Medicago.  
REFERENCE 1 (bases 1 to 1618)  
AUTHORS Fahrendorf, T., Ni, W., Shorosh, B.S. and Dixon, R.A.  
TITLE Stress responses in alfalfa (Medicago sativa L.) xix.  
TRANSCRIPTIONAL activation of oxidative pentose phosphate pathway  
genes at the onset of the isoflavonoid phytoalexin response  
JOURNAL Plant Mol. Biol. 28 (5), 885-900 (1995)  
MEDLINE 95367649  
REFERENCE 2 (bases 1 to 1618)  
AUTHORS Fahrendorf, T.  
TITLE Direct Submission  
JOURNAL Submitted (06-DEC-1994) Theo Fahrendorf, Samuel Roberts Noble  
Foundation, Plant Biology Division, 2510 Sam Noble Parkway,  
Ardmore, OK 73402, USA  
FEATURES  
source  
1. .1618  
Location/Qualifiers  
/organism="Medicago sativa subsp. sativa"  
/cultivar="Apollo"  
/sub\_species="sativa"  
/db\_xref="taxon:56147"  
/clone="019"  
/tissue\_type="cell suspension culture"  
2. .1467  
/codon\_start=1  
/product="6-phosphogluconate dehydrogenase"  
/protein\_id="AAB41553.1"  
/db\_xref="GI:603221"  
/translation="MAPPLTRIGLAGLVNQMALNIADKPPISVYNTTSKVDET  
VERAKGNPLFGFHPDFANVSIKPRVIMLVKAGAPVETTKLSAYLEKGC  
LLKVAQVPPGCVTVVGGSGNFKVIMHNGIEYGMQIAEAYDLVLSVGLTNE  
ELOSFTFNKGLLSFLIETADIFGDKDGDVLDKVDKTKMGKTKVTVQQA  
AELSVPAETASLDARLSGLDERVEAAKFKSGFGDILTDQOVKKQLDDVVRK  
ALYAAKICSTYAGQNMNLIRAKSAEKGDLALGELARIKWKGCIIIRAFIDRIKQAYRN  
PNLANLLVDFEFAKEIIRTOTAMRVVSLVNSGISLPGMSASLAYFDSTRRLRFPAN  
LVQQRDYFGAHTYVDIEGSHTEWFKLAKOSRI"  
1618  
polya\_site /note="24 A nucleotides"  
BASE COUNT 461 a 284 c 405 g 468 t  
ORIGIN  
Query Match 52.7%; Score 216.2; DB 45; Length 1618;  
Best Local Similarity 70.7%; Pred. No. 6.2e-35;  
Matches 287; Conservative 0; Mismatches 119; Indels 0; Gaps 0;  
QY 5 cgcgtccgcgttcagtgaggagacactccaagtgagcagcgtgacgctgccaagcgaga 124  
DB 39 CGGATTGGCTGTATGGGCGAGAATCTCGCACTCAATATTCGCGACAAGGATCCCAAT 98  
QY 65 ctctgtgtaacagagacacactccaagtgagcagcgtgacgctgccaagcgaga 124  
DB 99 TTTCTGTTTATATAGAACCAACATCAAGGTTGACCAACTGTGGACGACCAACAGGA 158  
QY 125 agaaacacctccctcagcttcacgtccatgaccgcccgctccttctgtgaagtcattcaga 184  
DB 159 GGGAAATCTTCACACTCTATGGCTTCCATGACCTGAAGCCTTTGTATATTCATTGA 218  
QY 185 gccacgggtggtgatcatctcgtcaagcgccgacgagtcaccagcagcagcagcagc 244  
DB 219 ACCAAGGGTGATTAAATGCTGTGTTAAGGCTGGTGCACCTGTGACACACACCAAC 278

QY 245 gctcgcagctcacttgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 304  
DB 279 CCTATCTCGTACCTTGAAAAGGTTGATTGATATATTAATGATGATGATGATGATGATG 338  
QY 305 gaacacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 364  
DB 339 GAACACTGAGAGCAG 398  
QY 365 tctctctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 410  
DB 399 AGTTTTCAGTGTGAG 444  
RESULT 10  
LD130772 3158 bp mRNA PLN 17-NOV-1999  
LOCUS Laminaria digitata mRNA 6-phosphogluconate dehydrogenase.  
DEFINITION  
ACCESSION AJ130772  
VERSION AJ130772.1 GI:6453556  
KEYWORDS 6-phosphogluconate dehydrogenase; 6PGDH gene.  
SOURCE Laminaria digitata.  
ORGANISM Laminaria digitata.  
Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group;  
Phaeophyceae; Laminariales; Laminariaceae; Laminaria.  
REFERENCE 1 (bases 1 to 3158)  
AUTHORS Crepineau, F., Roscoe, T., Kaas, R., Kloareg, B. and Boyen, C.  
TITLE Alternation of generations in laminaria digitata: a comparison of  
the gametophyte and sporophyte physiology by an EST strategy  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3158)  
AUTHORS Moulin, P., Crepineau, F., Kloareg, B. and Boyen, C.  
TITLE Isolation and characterization of six cDNAs involved in carbohyrate  
biosynthesis in Laminaria digitata (Phaeophyta) - Utilization of an  
EST database  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 3158)  
AUTHORS Boyen, C.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-1998) Boyen C., Laboratoire de Biologie  
Moleculaire et Cellulaire des Algues, Station Biologique - CNRS UPR  
9042, BP 74, 29682 Roscoff cedex, FRANCE  
FEATURES  
source  
1. .3158  
/organism="Laminaria digitata"  
/db\_xref="taxon:80365"  
/dev\_stage="sporophyte"  
68. .1660  
/gene="6PGDH"  
68. .1660  
/gene="6PGDH"  
/EC\_number="1.1.1.44"  
/codon\_start=1  
/product="6-phosphogluconate dehydrogenase"  
/protein\_id="CA61332.1"  
/db\_xref="GI:6453557"  
/translation="MSASALSDVGLYGLAVMGQNFALNWSHGFSVSNRSPEKVIDA  
TVARAKEGNLPLRGFKDPKSFVDSLSKPKIVLLVQAGAADVATLTLSELLEGGDI  
LVVDGNEFPNSQRRAKLEPKKIMFVGVSGSGEGEAGNPSLMPGGPREAYDALSP  
ILSKCAAOVDGPGCTTVVLSIGSVNKMVHNGIEYGMQIAEAYDLVLSVGLTNE  
LSNTFTDNGELOSFLIETAVIFSRKDLTSSGVVDKILDKTGMKGTGRTVQEA  
AERSTAAPMAALADARYMSGRGERVEAAKTLKGPSTIPQVPAQLIDEMKAAALYCS  
KICSYAQGMNLIRAKSAEHEFWGVDLGECAIRWKGCIIIRAFIDRIKAYIKNADLDN  
LVDPFITTSQILERQAARVRVTLVCAAPAMTASLUSYFDSYFDSYFDSYFDSYFDSYFDSYF  
RPFPGATYRVDKEGVFHLWDTHKEIATSPAHGRTLLIGSLNPLRPLCMGCLNW  
SSVVSVAVCGMKYVIGLSLFT"  
BASE COUNT 686 a 906 c 936 g 630 t  
ORIGIN  
Query Match 47.6%; Score 195.2; DB 8; Length 3158;  
Best Local Similarity 67.2%; Pred. No. 1.1e-30;  
Matches 275; Conservative 0; Mismatches 134; Indels 0; Gaps 0;



RSKFDKSFSCSSSLNCGDVTYTPVMDKLEERTFTVTYSIVTPLDNNAFQVATLAD  
 KGCIVQLSDSCCTGCDGORYPCPLHALAVCKKLKLPNLOVDDCYTLERKLRYAT  
 IFSHVPEMSAWPAGVRLPPVPPSPPTTYSVREKVMWVCHLNGITKIGAD  
 KYVEGSIILKTIIMKTAISLDRLYLQFLDKQSEKIFGYPVDVSDPDLTYV  
 HPVLNDSLETWLEVPNHPVSNVLEFLQAQPTSDAVVPTVCSPLSPDSSKR  
 ORTOOEIVDNNSSGAILVKNVSGALKPCLLPLRIWDDHMHGLGCFKDRDELKK  
 AWDWCIIRBRNCIVRETEKEMVTECVKWKWSLRAARMEHGLVETTKYTGPHTC  
 SHYPNDPSEFADEIERVRIQPTLSIAELKWKWKETGYELQTSKMRDGLKLVK  
 RVGEDQSFVMPKLIISAFSHSNGLLVDQDLPNDFPSFGRVFWFSQSIQEGFQ  
 HCRPLIVPTKSLNGYOLKLMIAAGVADKEFFPLAFATKEVTSDSNRWFFTKIRE  
 KTVQRDLCLISLPDIIVAVNPEGSLWQEPWAHFKCLNHLRSQFLGVRFDYNLES  
 LVQAGCTNOKBEFDSYMDIIEKNPEAKWLDIOIPRHKWALAHSGRLYPIEIDRE  
 ALFACVGPCTVACTVGMVMDDELRSDFKSLSIYSSLNRGVWYTYEPIEDLEE  
 FMTDPSIVITQLERSFVSSSEKEEINVLNVSTCTCRKFQSKFKFCLHALAVEE  
 KLINPQVDECTVEQKYTAATFSPVDPVAAWPCDRCVPTFSPFQSQSPNP  
 complement(join(15674..15817,15904..16008,16075..16196,  
 17033..17102))

/note="similar to ubiquitin-conjugating enzyme E2-17 kd  
 spiP35135; similar TO ESTs gb|184048, gb|745251,  
 gb|A100555, gb|AA041165, gb|T21697, emb|429145,  
 gb|T20867"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F22C12.2"  
 /protein\_id="AAF24583.1"  
 /db\_xref="GI:6692118"  
 /translation="MASKRLIKELKLDKQDPTSCSAVAEDMFWQATIMGPSDSPYS  
 GGVLVIHPPDPYPPKPPKVPATKVFHPNVSNGISCLDILKEQWSPALTSKVVLL  
 SIGSLTDPNPDPLVPEAHMYTKDRAKYTESARSWTQKYAM"  
 complement(join(18031..18617,18685..19861))  
 /note="putative receptor kinase emb|CAA23040.1"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F22C12.3"  
 /protein\_id="AAF24582.1"  
 /db\_xref="GI:6692117"

/translation="MOITLFFSLILCFVLISQTLDDKALLHLFSSFSRSLHN  
 QSDVCHSVGTCNENGRDVSIVRPAVGNGLIPPTISRLSLKFLSLRNHFTG  
 DFPDFTNLSKVLQHNHLSQPLIAIFSELKLNKVLDDLSNNGPSTPISLGLT  
 SLQVNLANSFSGEIPNLHLPKLSQINLNKLGITIPKSLQRFSSAFSGNNLTER  
 KQKRTFPGSLUAFLLISAACVLCVSLGFIWTCFKTRISGRLKRDSSPPGN  
 WTSRDDTEGGKIIFGGRNHLFDLDDLSAASEVLGAGFTTYKVMEDMSTVVV  
 KRKEVYVRREFEQMEIKMIRHENVAELKAYYSKDDKLVAYYHNGSFEILH  
 GNRGYHRVLDARUATGAARGLAKTHEGKFTGNKKNISFIYDSCYICIGDV  
 GLTIMSLPQTLCTSCYHAPETDTRRSTQFSDVYSGVLELLTGTGSPVSOAL  
 VPTGNNMLDIASIRSVVAKETGWFVDMELLSQSGGFEEMVLEIQLGLACVALQKQ  
 ERPHIAQVLKIEDIRSDAE"  
 join(20892..20993,21210..21310,21520..21694,21780..21976,  
 22063..22162,22282..22387,23017..23183)  
 /note="similar to vacuolar ATPase gl|3600058"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F22C12.4"  
 /protein\_id="AAF24559.1"  
 /db\_xref="GI:6692094"  
 /translation="MNDADASIQQQVRFIROAEAEKANEISISEEENFIEKQLV  
 EAEKKIROEYKKEKQVDRKKIDYSMOLNASRIKVLQADDDIVNMAEAAKQLLK  
 VSOQGFNHHHXYKHLKDLVQCLIKPEPAVLLRCEDDIDIVESMLDDASEVC  
 KKAHAPETIIVDKDIFLPAPSDDDPHALSCAGGVVLAARDKIVGENTILDLARLVA  
 FRNKLPEFCSGSFLEKVDKVALRGQWSLMSDSNFIETKRLDRAKSNMPTGRKCP  
 DPNGVEKSMCYSSCKTQGFMGSCQGHKNYMECEY"  
 23882..25345  
 /note="putative 6-phosphogluconate dehydrogenase  
 dbj|BA22812; similar to ESTs emb|F15308, emb|15307,  
 gb|T21894, gb|A199770.1"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F22C12.5"  
 /protein\_id="AAF24560.1"  
 /db\_xref="GI:6692095"  
 /translation="MESALSRIAGLAVMGONALNIAFKGPIPVNRTTSKYDE  
 TLDRAAEGNLPVSGQSPDRFVLSIQRSLIIVKAGAPVDOTIDAFSEYRPGDC  
 IIDGNGENYQNTERRISEAQKGLLILGMGVSGEGEAGRPGLMPGSGQAYDNKID

ILEKVAQVEDPCVTVIGGGGNGFVKVHNGITEYGMOLISEAYDVEKNVGLSNE  
 ELAIFTEWNSGELSEFLVITSDIFRVKDFGDELVDKILDKTGKGTGKWTVDQQA  
 AELSVRAPTAASDCRYLSGLDERENAAKVLREAGLKEIGSASSGIDKRLVDDV  
 ROALYAKISYAGCMNLLRAKSLKSWNLNFGELARIWKGCCIIIRAVFLDRIKKAYQ  
 ANLVQASLVVDPEFAKEMVQRAARVVRGLAVSAGISITPMCASLAFETYRRARP  
 RNLVQARQDLFGAHTYERTDPGAYHTWTKLARKNH"  
 complement(join(25806..26028,26204..26373,26722..28161,  
 28379..28489,28868..29254))  
 /note="hypothetical protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F22C12.6"

Query Match 44.6%; Score 183; DB 7; Length 111222;  
 Best Local Similarity 66.0%; Pred No. 1.5e-28;  
 Matches 264; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
 QY 11 CQCGGTGAGGCGGAGAACCTTCCCTCAACATGCGAGAGAGGTTCCCATCTCTGT 70  
 DB 23923 CGCGGTGAGGCGGAGAACCTTCCCTCAACATGCGAGAGAGGTTCCCATCTCTGT 23982  
 QY 71 gtacacagagcaacctccacaggtgagagagaccgtgagagagagagaa 130  
 DB 23983 CTACAATGAACACCTTCCAAAGTCGAGAAACCTAGATCGCGCGCTCGAAGGAA 24042  
 QY 131 cctcccgctacgcttcacatgaccccgctcctttgtgaagtcattcagagagccagc 190  
 DB 24043 CTTCCCGATCTCGGCCAATATCTCTCCCGGATTCCTCTCGATCCACAGGCTAG 24102  
 QY 191 ggtggtgatcgtcgtcgaagccgagccgagccagtcagacacacacgacgtcgc 250  
 DB 24103 ATCTCTCATCTCTCGTCAAAAGCTGGAGCTCCGTTGACCAAAACATCGAGCGCTTCTC 24162  
 QY 251 agtcaactggagcagcagcactgcacatcatcagtgagggagacagtgatcagagaaac 310  
 DB 24163 CGAATACATGGAGCCAGGAGATGTCATCATCGGTTGGAACAGATGGTATCAGAACAC 24222  
 QY 311 ggagagagggagagagccatgagagcgcgccctctgtatcttggcagtggtgtctc 370  
 DB 24223 AGAGCGACGATCTCGAAAGCTGACGAAAGAGATTCCTATTAGTATGGAGTCTC 24282  
 QY 371 tggagagagggaggtgcccgaacgagccgctccttgatg 410  
 DB 24283 CGCGCGGAGAGAGGAGCTCGTAACGGTCTCTCGTTAATG 24322

## RESULT 12

AB005233/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

AB005233 79837 bp DNA PLN 20-NOV-1999  
 Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MBK23,  
 complete sequence.

AB005233.1 GI:2264305  
 HTG.  
 Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui P1  
 clone:MBK23.

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.

1 (sites)  
 Sato, S., Kotani, H., Nakamura, Y., Kaneko, T., Asamizu, E., Fukami, M.,  
 Miyajima, N. and Tabata, S.

Structural analysis of Arabidopsis thaliana chromosome 5. I.  
 Sequence features of the 1.6 Mb regions covered by twenty  
 physically assigned P1 clones

DNA Res. 4 (3), 215-230 (1997)  
 97471969  
 2 (bases 1 to 79837)

Direct Submission  
 Submitted (02-JUL-1997) to the DDBJ/EMBL/GenBank databases.

**TITLE**  
JOURNAL  
Direct Submission  
Submitted (28-APR-2000) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA







KEYWORDS HTG; HRGS-PHASE1.  
SOURCE Leishmania major.  
ORGANISM Leishmania major.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
REFERENCE 1 (bases 1 to 42000)  
AUTHORS Myler, P.J., Sisk, E., Cawthra, J., Handley, F., Vogt, C., Robertson, L., McDonagh, P., Ivens, A., Nguyen, D., Munden, H. and Stuart, K.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUN-2000) Seattle Biomedical Research Institution, 4 Nickerson Street, Seattle, WA 98109-1651, USA  
COMMENT \* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 1990: contig of 1990 bp in length  
\* 1991 2023: gap of unknown length  
\* 2024 3664: contig of 1641 bp in length  
\* 3665 3697: gap of unknown length  
\* 3698 6805: contig of 3108 bp in length  
\* 6806 6838: gap of unknown length  
\* 6839 10632: contig of 3794 bp in length  
\* 10633 10664: gap of unknown length  
\* 10665 13449: contig of 4785 bp in length  
\* 13450 15482: gap of unknown length  
\* 15483 19169: contig of 3687 bp in length  
\* 19170 19200: gap of unknown length  
\* 19201 22797: contig of 3597 bp in length  
\* 22798 22829: gap of unknown length  
\* 22830 27115: contig of 4286 bp in length  
\* 27116 27147: gap of unknown length  
\* 27148 32918: contig of 5771 bp in length  
\* 32919 32950: gap of unknown length  
\* 32951 42000: contig of 9050 bp in length.  
FEATURES  
source  
1. 42000  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/chromosome="35"  
/clone="L5024"  
BASE COUNT 8701 a 11999 c 11887 g 9113 t 300 others  
ORIGIN  
Query Match 26.9%; Score 110.4; DB 57; Length 42000;  
Best Local Similarity 56.4%; Pred. No. 1.3e-13;  
Matches 226; Conservative 0; Mismatches 172; Indels 3; Gaps 1;  
QY 9 tccgcgtcatggggcgaacacctgcctcaacattgcagagaagggttccccatctct 68  
Db TCGCGGTACCTGGCGGCGGATCTCGCCTGAACATCGCCGAGAGGATTTAAAGTTGCC 7530  
QY 69 gtgtacaacagacaaacctcaagtggacag---acctgacgtgccaaggcagaa 125  
Db GTCTTACCCGACCTAGCGGAGAACGACGCTGCTTCTCAAGGAGCATGAGCGGAGAA 7470  
QY 126 ggaacaccttccgtctacggttcctatgacccgcgtctcttctgtgaagtcattcagaag 185  
Db TTTGCCGCCAACCTGAATGATAGGACCATGAGAGAGTTTCCTGCTCCCTCAAGAAG 7410  
QY 186 ccaggtgtgtatcatgctgtctgaagccggtgcgagtgaccagaccatcgagcag 245  
Db CCGCGCGCGGTTCATCTCTGTCGTCGACGCGCGCGCTACTGACTCTACAATCGAGCAG 7350  
QY 246 ctgcagctcacttgagcagggcgactgcattcatcgatgggggaacagtggttcagag 305  
Db CTCAAGGAAGTTCGAGAACGGCGACATCATATCGACACTGGTAATCGGAACCTTCAAG 7290

QY 306 aacacggagagaggagagccatggagagcgcgccctnctgtatcttggcatgggt 365  
Db GACACGACGAGCGCGGCTCAGTTTGAGAGCGAGGCTCCTCGCTTCCTCGCATGGGC 7230  
QY 366 gtcttgagagaaaggaggtgcccgaacagcgccctccct 406  
Db ATCTCTGTGTGAGAGGGTGGCGCAAGGGCGCGCCTT 7189  
RESULT 15  
D89161  
LOCUS D89161 1683 bp mRNA PUB 13-MAR-1998  
DEFINITION Schizosaccharomyces pombe mRNA, partial cds, clone: SY 0651.  
ACCESSION D89161  
VERSION D89161.1 GI:1749529  
KEYWORDS  
SOURCE Schizosaccharomyces pombe (strain:PR745) cDNA to mRNA, clone:lib:library of H. Nojima clone:SY 0651.  
ORGANISM Schizosaccharomyces pombe  
Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomycetes.  
REFERENCE 1 (bases 1 to 1683)  
AUTHORS Yoshioka, S.  
TITLE Direct Submission  
JOURNAL Submitted (15-NOV-1996) to the DDBJ/EMBL/GenBank databases. Sachiyo Yoshioka, Tsukita Cell Axis Project ERATO JST, Kyoto Research Park; 17 Chudouji Minamimachi, Shimokyo-ku, Kyoto, Kyoto 600, Japan (E-mail:syoshi@cell.tsukita.jst.go.jp, Tel:+81-75-315-7913, Fax:+81-75-315-6420)  
REFERENCE 2 (sites)  
AUTHORS Yoshioka, S., Kato, K., Nakai, K., Okayama, H. and Nojima, H.  
TITLE Identification of open reading frames in Schizosaccharomycetes pombe CDNAS  
JOURNAL DNA Res. 4 (6), 363-369 (1997)  
MEDLINE 98162722  
FEATURES  
Location/Qualifiers  
1. 1683  
/organism="Schizosaccharomyces pombe"  
/strain="PR745"  
/db\_xref="taxon:4896"  
/clone="SY 0651"  
/clone\_lib="library of H. Nojima"  
CDS  
/note="similar to Saccharomyces cerevisiae 6-phosphogluconate dehydrogenase(decarboxylating), SWISS-PROT Accession Number P38720"  
/codon\_start=3  
/protein\_id="BAA13823.1"  
/db\_xref="GI:1749530"

/translation="PTNQTFINHTTKRRIMSKQEVADFLIGLAVMGONLILNGADK  
GFTVCYNRRTSRVDFLANEAKGSIIVGAHSLSEFVSKLKKPRVCILLVAGKPVY  
LIEGLAPLLEKGGDIIVDGNNSHYDDTTRRCRELAKKGTILFVSGVSGEGEARGPSL  
MPGPNPAWPRKIPFOFLAAKAGNNEPCCDWGEQAGHVYKVVHNGIEYGDMLIC  
ETVYDLMKELGNSCDESDTVEKNTKIDSLFLEITRDVLRKADDDKPLVEKILDA  
AQCGTGTAGNALEMPVSLTEAVFARCLSKSERVRSKLTGPTKTKTGDK  
KOLIDLEDLALYASKIISYAQGFMLREAAKEYGKLNAGIALMRGGCIIIRSVFLK  
DITEAFREDPNLESLIFPFTNGVERAQAQWRVVAQAAMGLIPVPATSTGLSFYDG  
YRSVAVPANLLQAQRDYFGAFTFVLPADKSLPADKDIHINWTHGGNISATYDA  
BASE COUNT 397 a 359 c 398 g 529 t  
ORIGIN  
Query Match 26.7%; Score 109.4; DB 8; Length 1683;  
Best Local Similarity 56.2%; Pred. No. 4.4e-13;  
Matches 228; Conservative 0; Mismatches 172; Indels 6; Gaps 1;  
QY 5 cgcgcctccggttcattggggcagaaccttgcctcaacattgcagagaagggttcccat 64  
Db CGGTTTGGCGGTGATGGGTCAAAACTTGTTCACCGTGCAGCAAGGCGCTTACCGT 145  
QY 65 ctctgtgtacaacagacacacctccaagtgagcagacctgcagcgtgcgaaggcaga 124

Db	146	CGTGTGCTACACCGTACAACTTCAGAGTTTACGAGTTTGTG-----GCAACGAAGC	199
QY	125	aggaacacctccgctctacggtctccatgaccccgctcccttggtagatccattcagaa	184
Db	200	CRAAGGTAATCTATTGTGGTGCTCACTCCCTTGAGGAATTTGTTCCAGGCTAAGAA	259
QY	185	gcacgggtggtgatactgctcaagcgcgccagtgctgaccagacatcgac	244
Db	260	GCCTCGTGTGTATTTCGTTGTTAAGGCTGGTAAGCTGTGTATTACCTATTGAGGG	319
QY	245	gctcgagctcaattggagcaggcgactgcatcatcgatgggggaacagagtggtaga	304
Db	320	ATTGGCTCTCTCCTCGAGAAGGGGATATCAATTGTACGGTGTGTAATCCCATACCC	379
QY	305	gaacacggagaggaggaagggccatggagcgcgccctnctgtatcttggcatggg	364
Db	380	TGATACTACCCCGCTGCGAGGAGTTGCCAAGAAGGCGATCTTTTGTGTGTTCCGG	439
QY	365	tgctctgaggaaaggagggtgcccgcaacggcccgctctgtagc	410
Db	440	TGTTTTCTGGTGTGAAGAGGGTGCTCGTTACGGTCCCTCTTTGATG	485

Search completed: November 4, 2000, 13:32:00  
Job time: 16969 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:40:13 ; Search time 4075.18 Seconds  
(without alignments)  
321.525 Million cell updates/sec

Title: US-09-300-482-619  
Perfect score: 300

Sequence: 1 gattattgacaacaccgat.....ttcagaaatgtctgtgtgg 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_bal:\*

2: gb\_ba2:\*

3: gb\_om:\*

4: gb\_ov:\*

5: gb\_pat:\*

6: gb\_ph:\*

7: gb\_pl1:\*

8: gb\_pl2:\*

9: gb\_pr1:\*

10: gb\_pr2:\*

11: gb\_pr3:\*

12: em\_fun:\*

13: em\_hum1:\*

14: em\_hum2:\*

15: em\_in:\*

16: em\_om:\*

17: em\_or:\*

18: em\_ov:\*

19: em\_pat:\*

20: em\_ph:\*

21: em\_pl:\*

22: em\_ro:\*

23: em\_sts:\*

24: em\_sy:\*

25: em\_un:\*

26: em\_vi:\*

27: gb\_htg1:\*

28: gb\_htg2:\*

29: gb\_in1:\*

30: gb\_in2:\*

31: em\_bal:\*

32: em\_ba2:\*

33: em\_hum3:\*

34: em\_hum4:\*

35: gb\_pr4:\*

36: gb\_htg3:\*

37: gb\_htg4:\*

38: gb\_htg5:\*

39: gb\_htg6:\*

40: gb\_htg7:\*

41: em\_htg1:\*

42: em\_htg2:\*

43: em\_htg3:\*

44: em\_hum5:\*

45: gb\_pl3:\*

46: gb\_pr5:\*

47: gb\_htg8:\*

48: gb\_htg9:\*

49: gb\_htg10:\*

50: gb\_htg11:\*

51: gb\_htg12:\*

52: gb\_htg13:\*

53: gb\_htg14:\*

54: gb\_in3:\*

55: gb\_htg15:\*

56: gb\_htg16:\*

57: gb\_htg17:\*

58: em\_htg4:\*

59: em\_htg5:\*

60: em\_htg6:\*

61: em\_htg7:\*

62: em\_hum6:\*

63: gb\_htg18:\*

64: gb\_htg19:\*

65: gb\_ba3:\*

66: em\_htg8:\*

67: em\_htg9:\*

68: em\_htg10:\*

69: em\_htg11:\*

70: em\_htg12:\*

71: em\_htg13:\*

72: em\_htg14:\*

73: em\_htg15:\*

74: em\_htg16:\*

75: em\_htg17:\*

76: em\_htg18:\*

77: em\_htg19:\*

78: em\_htg20:\*

79: em\_htg21:\*

80: em\_htg22:\*

81: em\_htg23:\*

82: gb\_pr6:\*

83: gb\_pr7:\*

84: gb\_htg20:\*

85: gb\_htg21:\*

86: gb\_htg22:\*

87: gb\_htg23:\*

88: gb\_ro:\*

89: gb\_sts1:\*

90: gb\_sts2:\*

91: gb\_sy:\*

92: gb\_un:\*

93: gb\_vil:\*

94: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description	
		Match						
1	199.2	66.4		2142	45	SOJ000265	AJ0000265 Spinacia	
2	79.2	26.4		4176	7	AF120494	AF120494 Arabidops	
3	79.2	26.4		125803	8	ATF22K18	AL035356 Arabidops	
c	4	79.2		26.4	198402	8	ATCHRIV61	AL161561 Arabidops
5	64	21.3		2792	65	SYCCYSEPGI	D13777 Synechocyst	
6	64	21.3		128598	2	D90912	D90912 Synechocyst	
c	7	36.2		12.1	15727	1	AE001224	AE001224 Treponema
c	8	35.6		11.9	156995	85	AL365270	AL365270 Homo sapi
c	9	35.4		11.8	101370	64	AL353604	AL353604 Homo sapi
c	10	35		11.7	11836	1	AE002356	AE002356 Chlamydia
c	11	35		11.7	162149	48	AC023104	AC023104 Homo sapi
c	12	34.6		11.5	33328	40	AC019951	AC019951 Drosophil

13 34.6 11.5 148447 36 AC010575 Drosophil  
14 34.6 11.5 207585 57 AC073779 Mus muscu  
15 34.6 11.5 235066 57 AC074208 Mus muscu  
16 34.6 11.5 319765 29 AE003481 Drosophil  
17 34.4 11.5 136570 39 AC016555 Homo sapi  
18 34.4 11.5 175191 10 AC010251 Homo sapi  
19 34.4 11.5 179667 50 AC025681 Homo sapi  
20 33.6 11.2 155069 51 AC027133 Homo sapi  
21 33.4 11.1 165921 40 AC019354 Homo sapi  
22 33.4 11.1 167271 27 AC006269 Homo sapi  
23 33.4 11.1 169140 52 AC036229 Homo sapi  
24 33.2 11.1 16879 38 AC015095 Drosophil  
25 33.2 11.1 152585 27 AC008025 Homo sapi  
26 33.2 11.1 152667 52 AC027702 Homo sapi  
27 33.2 11.1 158467 85 AL365446 Homo sapi  
28 33.2 11.1 164107 38 AC016012 Homo sapi  
29 33.2 11.1 164865 37 AC012483 Homo sapi  
30 33.2 11.1 167349 39 AC016099 Homo sapi  
31 33.2 11.1 168960 39 AC016100 Homo sapi  
32 33.2 11.1 203924 37 AC012159 Drosophil  
33 33.2 11.1 301929 29 AE003502 Drosophil  
34 33 11.0 149528 87 HS367B10 Homo sapi  
35 33 11.0 153140 52 AC034202 Homo sapi  
36 33 11.0 157609 50 AC025466 Homo sapi  
37 33 11.0 165334 28 AC008544 Homo sapi  
38 33 11.0 189925 87 CNS01DX4 Homo sapi  
39 33 11.0 190685 49 AC023587 Homo sapi  
40 32.8 10.9 18765 1 AE000961 Arabidopsi  
41 32.8 10.9 19590 87 PFMA113P5 Arabidopsi  
42 32.6 10.9 137502 64 AL161737 Homo sapi  
43 32.6 10.9 120180 63 AF279873 Homo sapi  
44 32.6 10.9 126474 46 HS181N1 Human DNA s  
45 32.6 10.9 135404 63 AF228658 Homo sapi

ALIGNMENTS

RESULT 1  
SOJ00265 2142 bp mRNA PLN 25-AUG-1998  
LOCUS Spinacia oleracea mRNA (nuclear-encoded) for chloroplast  
DEFINITION glucose-6-phosphate isomerase.  
ACCESSION AJ000265  
VERSION AJ000265.1 GI:3413510  
KEYWORDS glucose-6-phosphate isomerase.  
SOURCE Spinach.  
ORGANISM Spinacia oleracea  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
AUTHORS Euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
TITLE Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
JOURNAL 1 (bases 1 to 2142)  
MARTIN, W.F.  
Direct Submission  
Submitted (16-DEC-1997) Martin W.F., Spielmannstr. 7, Institut fuer  
Genetik, TU Braunschweig, D-38023 Braunschweig, D-38023, GERMANY  
REFERENCE 2 (bases 1 to 2142)  
AUTHORS Nowitzki, U., Flechner, A., Kellermann, J., Hasegawa, M.,  
Schnarrenberger, C. and Martin, W.  
TITLE Eubacterial origin of nuclear genes for chloroplast and cytosolic  
glucose-6-phosphate isomerase from spinach: sampling eubacterial  
gene diversity in eukaryotic chromosomes through symbiosis  
JOURNAL Gene 214 (1-2), 205-213 (1998)  
MEDLINE 98322268  
FEATURES Location/Qualifiers  
source  
1. .2142  
/organism="Spinacia oleracea"  
/db\_xref="taxon:3562"  
/note="nuclear gene for chloroplast enzyme"  
transit\_peptide 84..263  
/gene="GPIP"  
/product="nuclear encoded chloroplast GPI"  
84..1940  
CDS

/gene="GPIP"  
/EC\_number="5.3.1.9"  
/function="glycolysis, gluconeogenesis"  
/note="nuclear encoded chloroplast GPI"  
/codon\_start=1  
/product="glucose-6-phosphate isomerase"  
/protein\_id="CAA03982.1"  
/db\_xref="GI:3413511"  
/db\_xref="SPTREMBL:O82058"  
/translation="MASSLSNLYSTSLKPKISHLPTITKNPISPKSLSPKPISSVA  
RDTPADLSTSSSSNNLPSLQKKADSLKEDPRALMARVEMLYOHKDLGLVDVS  
RIGFDEFVKEMEFKAFKMEGIEKGAIANDEGRMWHYWLRLSSSLAPPTFLKN  
QIDVLDRWQFANDVIGSKIKAPTGERFTHILSVGIGGSALGPOFAEALAPNDPL  
KTRFDNDTPAGIDHIOALGPDLATLVMVISKGGTPTNRGLLEVKAFQADGLV  
FAKQGVALTQENSLDNTARTEGMDRPFMDWGGRTSEMSAVGLLPAALQGLDIKE  
MLAGALMDEATKIPVLSNPAALLANSWYASDGVSKDMVLPYKDSLLLSRYLQ  
QIVMSLEKFDLDKNVQGLTVYVGNKSGSDOHAYIQOLRGVHNFATFIEVLDR  
PFGHWELEPGVTGCDYLFGLFMGYSALYANNRESISVTVOEVTPTSRVGMALYER  
AVGLYASLVNTINAYHOPGVKAAAEVLALQKRVLAVLNEASCKDPVEPLTIEEVA  
DHCHCPDDIEMIYKIIAHMAANDRVILAEAGDCGSPRSIKAFIAGCNVDLYA"  
84..1940  
/gene="GPIP"  
mat\_peptide  
264..1937  
/gene="GPIP"  
/EC\_number="5.3.1.9"  
/note="nuclear encoded chloroplast GPI"  
/function="glycolysis, gluconeogenesis"  
/product="glucose-6-phosphate isomerase"  
BASE COUNT 593 a 467 c 490 g 592 t  
ORIGIN  
Query Match 66.4%; Score 199.2; DB 45; Length 2142;  
Best Local Similarity 79.0%; Pred. No. 8.8e-51;  
Matches 237; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
QY 1 gattattgacaacacgcgtctgctgggattgatacaaaattgctcaactagagcctg 60  
Db 745 GATTATTGATAATACAGACCCAGCTGGAATCGATCATCAAAATGACAACTTGGCTG 804  
QY 61 aactggcaactactctgttaattgcttaagagcggagcgcacacctgaacccgca 120  
Db 805 AGCTAGCTACACACTTGTATGTTTATTCARAGAGTGGCGGTACACCAACAAAGAA 864  
QY 121 atggtctactagaagtacagaagccttcagagatcggggctgcaattctcgaacagg 180  
Db 865 ATGTTTATTGGAAGTACAAAAGCGCTTCGGTGTGATTAGTAGTATTCGCAAAACAGG 924  
QY 181 gtgttgcaactactcaagaataattctctgttgatacaactgctagatagaggatggt 240  
Db 925 GTGTGCTATTAACACAGAAACTCATTTGTAGATATATCTGCACGAAATGAAGGCTGA 984  
QY 241 tagctcggttctctatgtttgattgggtggtggtggtggtggtggtggtggtggtggtg 300  
Db 985 TAGATAGATTTCCCATGTTTCGATTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTG 1044  
RESULT 2  
LOCUS AF120494 4176 bp DNA PLN 12-JAN-2000  
DEFINITION Arabidopsis thaliana phosphoglucose isomerase precursor (PGI) gene,  
complete cds; nuclear gene for chloroplast product.  
ACCESSION AF120494  
VERSION AF120494.1 GI:6690394  
KEYWORDS thale cress.  
SOURCE Arabidopsis thaliana  
ORGANISM Arabidopsis thaliana  
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
TITLE Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 4176)  
Yue, T.S., Yue, W.L., Wang, S.M. and Chen, J.  
Mutation of Arabidopsis chloroplastic phosphoglucose isomerase

affects starch synthesis and floral initiation

JOURNAL  
Unpublished  
2 (bases 1 to 4176)  
Yu, T.S., Lue, W.L., Wang, S.M. and Chen, J.  
Direct Submission  
Submitted (15-JAN-1999) Institute of Molecular Biology, Academia  
Sinica, Taipei, Taiwan 11529, Republic of China  
LOCATION/Qualifiers  
1. 4176  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="4"  
/map="67.3 cm"  
Join(338..856,1067..1186,1384..1482,1566..1649,  
1932..2076,2146..2210,2365..2433,2522..2629,2891..2943,  
3034..3172,3259..3415,3507..3585,3675..3741,3827..3961)  
/gene="PGI"  
/product="phosphoglucose isomerase precursor"  
/c38..>3961  
/gene="PGI"  
Join(338..856,1067..1186,1384..1482,1566..1649,1932..2076,  
2146..2210,2365..2433,2522..2629,2891..2943,3034..3172,  
3259..3415,3507..3585,3675..3741,3827..3961)  
/gene="PGI"  
/codon\_start=1  
/product="phosphoglucose isomerase precursor"  
/protein\_id="AAF24124.1"  
/db\_xref="GI:6690395"  
/translation="MASLGLVSSPSLKPKNHFKGTAGAIISRLFFPYLPHQS  
TLTLSSARSARDISUADSKKELLDPDALWKRYLDWFYQKELGLYLDISRGFTDE  
FVAEMERFOAAKAMEDLEKSGFIANDPDEGRMVGHYLRNSKLAPKPLKLIENLTD  
SICAFSDIIISGKIPSPPEGRFTQILSVGIGGALGQLSLRLGPDNPPLKIRFID  
NTDPAGIDHQAIGLPELSTLVVIVSKSGGTPETRNGLLEVQKAFREGLNFAKQV  
AITQENSLDNTARIEGLWLRPMYDWVGRSTIMSANGLLPAHLOGINVREMLTGA  
LMDEATRTTSIKNPAALLAMCWYASNGVSKDMVVLPIKDSLLIFRYLQQLVMS  
LGPEFDLGNVNOGLTVYNGKSTDOHYITQOLKRDGVHNFATFIEVLRYPGHDW  
LEPGFTCDGYLFGLQGRSALYANGRESISTVTOEVTPTSVGAIIALYERAVGLYA  
SIYNINAYHOPVGAAGKAAAEVLALQKRVLSVLEATKDPVEPLTLEETADCHAP  
EELEMIYKIIAHMSANDRVLIAEGNCGSPRSIKVYLGECNVDDLYA"  
338..556  
/gene="PGI"  
transit\_peptide 338..556  
polyA\_site 4109  
polyA\_site 4176  
BASE COUNT 1153 a 829 c 859 g 1335 t  
ORIGIN

Query Match 26.4%; Score 79.2; DB 7; Length 4176;  
Best Local Similarity 77.4%; Pred. No. 1e-13; 28; Indels 0; Gaps 0;  
Matches 96; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 177 cagggtgttcactcaagaaaattctctgttgataacactgctagaatagaggga 236  
|||||  
Db 1929 CAGGGTGTTCATAACCCAGAGAAGACTCATCTAGATAATACGCAAGATTGAAGT 1988  
|||||  
QY 237 tggtagctcggttctctattgttgattgggttggttgtagaacttcgaaatgctgct 296  
|||||  
Db 1989 TGCTAGCTAGATTCTCTATGTACGACTGGGTGGGGAAGAACAATCAATAATGCTGCA 2048  
|||||  
QY 297 gtgg 300  
|||  
Db 2049 GTTG 2052

RESULT 3  
ATF22K18 125803 bp DNA PLN 03-FEB-1999  
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone F22K18 (ESSAI  
project).  
ACCESSION AL035356  
VERSION AL035356.1 GI:4220510  
KEYWORDS thale cress.  
SOURCE

ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;  
Brassicaceae; Arabidopsids.  
REFERENCE  
AUTHORS Bevan, M., Wedler, H., Wedler, K., Wambutt, E., Hoheisel, J.,  
Mewes, H.W., Mayer, H.F.X. and Schueller, C.  
JOURNAL Unpublished  
2 (bases 1 to 125803)  
EU Arabidopsis sequencing, project.  
DIRECT SUBMISSION  
Submitted (03-FEB-1999) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:  
schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
E-mail: michael.bevan@bbsrc.ac.uk  
COMMENT  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of chromosome 4 can be  
viewed at: http://websvr.mips.biochem.mpg.de/proj/thal/  
FEATURES  
source  
1. 125803  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
171..2980  
/gene="F22K18.10"  
complement(join(171..601,689..883,983..1159,1250..1469,  
1556..2980))  
/gene="F22K18.10"  
complement(171..601)  
/gene="F22K18.10"  
/number=1  
complement(join(171..601,689..883,983..1159,1250..1469,  
1556..2980))  
/gene="F22K18.10"  
/note="similarity to DNA polymerase III gamma subunit -  
Aquifex aeolicus, PIR2:A70460  
Contains ATP/GTP-binding site motif A (P-loop) [PRTGTGT]"  
/codon\_start=1  
/product="DNA polymerase III like protein"  
/protein\_id="CAA22984.1"  
/db\_xref="GI:4220511"  
/translation="MGETRRHSVDVPIRTLVLRVRLRDPCTTSMKSPASLLDNV  
KWETGNSNGISLQFVHADDAACKAAADAPVGLIPEGSYSIMELESGLDLKLSKVI  
NVEGDACSRSSERSCDLSVGRDLACNAPSIHVEAGSGRTRHYTKLASVGE  
YGRSLGSPMNSTNHSYVYGDVDFDSQNRCCGITYCWSRTPRPNQSSDVEEYPL  
LPGNGGSDVVTPSHEVLSLSQKFRPKSFDLGVGVVVKLLSTILGRITTSYL  
LFHGPRTGKTSKIFAAALNCISQAHSRPGCLCECKSYFSGRGRDVTDSGL  
NRPYSRLSLKASLPVSSSRKVFIIDECQLCQETGTLNLSLDFSOHSFVLVT  
SELEKLPNRLSVRSQKHYFSKVCADISTKLAKICIEGIDFGQAVDFTASKDGL  
RDAETMLDLQLLGLKTRITSLAYKLIGVSDDELDDLILDLAMSDDTSNTVIRARELMR  
SKIDPMOLISOLANVIMDIAGNSQESSATRLRLTRHTSEEMQKRLNALKILSDA  
EKHLRAKNOTTLTVALLOLNTDSSFAIDENGRNQINDVELSTSSGCPGVK  
SDAEKQGRNENVEVSKVTDLCCSDSLNKLRLKRGRLTSLTVDKGVAIAELFV  
TPQHVAREKSKLLIADFQSVLGCNVEIQMLIVISACSPPKAKAAASFFGLFSCS  
RMLHKSYLETTTDDSCDAKEPAVTNLSRCQGNLVRARSVRSSANASSRMSQSQG  
DATSAMCTPHPPGGEK"  
complement(602..688)  
/number=1  
complement(689..883)  
/gene="F22K18.10"  
/number=2  
complement(884..982)  
/number=2  
complement(983..1159)  
/gene="F22K18.10"  
/number=3  
complement(1160..1249)  
/number=3  
complement(1250..1469)  
/gene="F22K18.10"









Db 1711 ATGCTCTGCGGAAACGAAGCAGTATTGAAGCCAGCGCTTCATTTGCCGATTATG 1770

QY 181 gtattgaattactcaagaaattctctgttgataaacactgctagataagaggggt 240

Db 1771 CGTGGCGGTGACCATGCGCCGACGATTAATTTCCAGCAGCGCCAGCAGCAATGGT 1830

QY 241 tagctcggtttctctactgtttggttggttggttagaactcaagaatgctgctgtg 300

Db 1831 TACAAGCTTTCCCATGCGAGTATGGGTGGTGGGAGAACCTCGAGTTATCGCGCGTGG 1890

RESULT 6

D90912 128598 bp DNA BCT 07-FEB-1999

LOCUS Synecocystis sp. PCC6803 complete genome, 14/27, 1719644-1848241.

DEFINITION D90912 AB001339

ACCESSION D90912.1 GI:1653228

VERSION 3-chlorobenzoate-3,4-dioxygenase; 30S ribosomal protein S1; 30S

KEYWORDS ribosomal protein S2; 50S ribosomal protein L34; 50S ribosomal

protein L9; ABC transporter; COB protein; DNA primase; GDP-mannose

pyrophosphorylase; GTP cyclohydrolase II; H+/Ca2+ exchanger; PetG

subunit of the cytochrome b6f complex;

UDP-N-acetylmuramoylalanyl-D-glutamyl-2,

6-diaminopimelate-D-alanyl-D-alanine ligase; acetyltransferase;

anthranilate synthase component I; aspartate carbamoyltransferase;

carbonic anhydrase; catabolite gene activator protein; cyclase;

cytochrome c ytm; cytochrome c oxidase folding protein; elongation

factor TS; extragenic suppressor SubB; fatty acid desaturase;

ferredoxin; glucosyl-6-phosphate isomerase; heme oxygenase;

hemolysin; hydrogenase expression/formaton protein HycP; inner

membrane protein; acetyltransferase; oxygen independent

coprophorhylogen III oxidase; polyA polymerase; regulatory components

of sensory transduction system; renin-binding protein; sensory

transduction histidine kinase; serine acetyltransferase; tRNA-Arg;

tRNA-Ile; tRNA-Ser; transposase; tyrococyte protein cthy28kD; AT103.

Synecocystis sp. (strain:PCC6803) DNA.

Synecocystis sp.

ORGANISM Bacteria; Cyanobacteria; Chroococcales; Synecocystis.

REFERENCE 1 (sites)

AUTHORS Kaneo,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,

Miyajima,N., Hirose,M., Sugita,M., Sasamoto,S., Kimura,T.,

Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,

Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,

Yamada,M., Yasuda,M. and Tabata,S.

Sequence analysis of the genome of the unicellular cyanobacterium

Synecocystis sp. strain PCC6803. II. Sequence determination of the

entire genome and assignment of potential protein-coding regions

DNA Res. 3 (3), 109-136 (1996)

97061201

2 (bases 1 to 128598)

Tabata,S.

Direct Submission

Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshi

Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure

2; 1523-3, Yanauchino, Kisarazu, Chiba 292, Japan

(E-mail: tabata@kazusa.or.jp. Tel: +81-438-52-3933,

Fax: +81-438-52-3934)

Potential protein coding regions were assigned on the basis of

similarity search of the ORFs and Genemark analysis.

FEATURES

Location/Qualifiers

1..128598

/organism="Synecocystis sp."

/strain="PCC6803"

/db\_xref="taxon:1143"

complement(1..933)

/note="ORF\_ID:s111263"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAA18144.1"

/db\_xref="GI:1653229"

CDS

/translation="MTARLPPTAVLSIGAAATMGLKGLAYAITGVSGLLSDALEST

VNLASIVAFWALSIATPADSEHPFHGSKAEYFSSGLEGAFFVAALGIGYSVERL

LSPRLDONALGIALAATAATLANGVTAVILMRAGKLNIALAGTQSHLMTDWTGVS

VVAVAILFVGTWELDLTALGAGVNLVGTGTHLLRETISLMDQSLPQAOLQATTS

FVYEDQGVRFHLLTROAGSQSFISFVHLVPGHWTVQRGHDLCEAIEATTAERTIG

SRVTTTHLEPLEDPKSWOHPDEFPSPAPLNDRKPN"

complement(997..1482)

/note="ORF\_ID:s111262"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAA18145.1"

/db\_xref="GI:1653230"

/translation="MLPLPIIANGKGFIRALENDGALAVYAPLEGYEGRYORRLRAN

GYASISLSARGLDGVEAYLMQVHGVRPAHLGKNKIAOAGVPIYFAQPIAGYQLENL

GYASKGLVILWLEGYILISQTEIQDLSLTKRVPKLKLVLEMGDRVFRWQPLDLCQA

A"

complement(1707..2363)

/gene="tsf"

complement(1707..2363)

/gene="tsf"

/note="ORF\_ID:s111261"

/codon\_start=1

/transl\_table=11

/product="elongation factor TS"

/protein\_id="BAA18146.1"

/db\_xref="GI:1653231"

/translation="MAEITAQVLKELREKTGAGMMDCKKALKENEGDLEKSTEWLRQK

GIASADKSGRTAAEGLVHSYIHFGRGVLVVENCTDFVARGDRFDLVNDVAMQI

AQPNYIVSVADIPQEMVAKEKEIMGRDDLGGKPKANIKKIVGRIDKKELSL

"

complement(2517..3326)

/gene="rps2"

complement(2517..3326)

/gene="rps2"

/note="ORF\_ID:s111260"

/codon\_start=1

/transl\_table=11

/product="30S ribosomal protein S2"

/protein\_id="BAA18147.1"

/db\_xref="GI:1653232"

/translation="MPVVSADLLSSEGVHFGHQTNRWPRMDQYIYARNGVHIDLV

QTAQMEDAYEYVRSTNGKFLFVGTKQAAGTISOEACRGAHVYVNOGWLMT

NNETTRKVRDLKLEALEASGIDRRGKKGSMRLRELKLOKYLGLKINMRKLPDV

VVYQRRHNAIHECQKIGIPIIAMLDTNCDPDVDVPIPAINDAIRSKLVGKLA

DAIYGRGQPDVSDYDEFEDEGLDGNLEVEAEVEEAEEVAATPEA"

complement(3551..3623)

/product="tRNA-Ile"

3809..4927

/note="ORF\_ID:s1r1336"

/codon\_start=1

/transl\_table=11

/product="H+/Ca2+ exchanger"

/protein\_id="BAA18148.1"

/db\_xref="GI:1653233"

/translation="MSTKSKIFLVLLVFCPLSFAAHWLGWETTTFVLAGLAIPLAA

FMGTATEIAVYIGNAGGLNATFGNATLILAYIALKEGLLEVVKATLTGSIIGNL

ILVMGFVFLGRLYKQNFOPLAARLNASTMNLGVVAIILPTALOYSTGVETVLQ

NLSVAVAVLIGVLLSVFSMGTHAVLYDVGVAENMEMFELGEDVSEPPPEEKP

NLWLWTGVLLVTLGVAVESLLVGSLEVATSGLTALTGTGVILPIIIGNAAEHATA

VTVAMKDKMDSVMYSGSLQIAFFVAPVIVGVAIQGPMOLNFPFELVAVLVAV

LIVNISDGTGSLNWLEGLILLATYAIVALAFFHPTLV"

complement(4954..5535)

/note="ORF\_ID:s111258"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAA18149.1"

/db\_xref="GI:1653234"

/translation="MIKNDIWRQAQAQGMIEPPEPKLMQIEGQPAISFGLSSFGYD

IRLSAADFRIFRHPGTVPDKNPNHNLNATOLHRDSDGSYFLLPAHSGYGLGVAVER

LAIPANVTVLCIGKSTVARAGIANTLTPGAGWGHLLTLEFNSSSADCRITANEGIV



```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC65461.1"
/db_xref="GI:3322762"
/translation="MSGHSKWATIKHAKGAADAKRGOLFTEKFIKEISVAARMAGGQPO
ANPRILAKRAANMPKDNIERAIKKGTCGELSSGVSEELVYEGYAGGVAVLVEVL
TNKNNRAANVRNLFNRNGNLGSGAGSVYFNKRGVIEDSEQVDEALMELALEAG
AEDIQNAGCVLTIVTVPCGTFETVLESQAQWESLSAGISWVPDYLALDEETARKVL
KNIDLEEDVDQAVYSNADISLVL"
complement(1655..3262)
/gene="Tp0475"
/CDS
complement(1655..3262)
/gene="Tp0475"
/notes="similar to GB:AE000783 percent identity: 46.56;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate isomerase (gpi)"
/protein_id="AAC65462.1"
/db_xref="GI:3322763"
/translation="MNRNLDCAAYARLQAIIRAPSLKTVLCGPEGIERRYCTDAG
AGLRVHYAAKTVNEILITLALAADEQELVAKYDALRAGAINTGKRVLHLTRLG
VQSSLASLCEVRDMHAFYTKERYCAFAQVHEGSLTSRGAFDTVOVIGIGGS
DUGPRALYLALEGWAQRQAVKMTHE ISNVDPDAAALVLSKPLETLTILVSKSGT
TLETSNELFVHVRQGLEPHTFQVAVTSETSPILANNPOYLASFYMDDFIGGRYSS
SSVCGAVVLTAFGPQVGFHFLSGAARADRAQOEDIRRNALLDALIGVYERTILGY
BHTAVLPYSQALAREPAHLOQLOMESNGKSVNRFGIPYTKTGPVIFGEPGNGQHSF
YQHLQGTSVVPLQFIAFQHSQGLQDPIIRGSTGQKRLANVAQIVAFARKEHADA
NKTSGRPSLLTAKALTPTOTIGALLAHFENKIMFQGFANWLNSPDQSGVQLGKTLA
QHILAGEVGVLRAYADFLDLAHPTC"
complement(3326..4672)
/gene="Tp0476"
/CDS
complement(3326..4672)
/gene="Tp0476"
/notes="similar to GB:L23147 SP:P38502 PID:349834 percent
identity: 43.61; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="acetate kinase (ack)"
/protein_id="AAC65463.1"
/db_xref="GI:3322764"
/translation="MILLTNCGSSSVKYQVYNWTERAVIAGVVERVVTQAGSVITHE
VHGENHVRSPCTHEAVQLIISTLDHAHGVISDMSLIKAVGRVHLHADREVKS
VWVPTDLEFRSVQDLGPHNPANIRGIEAAQAVIPSPHCAIMDTAHQTMPEASF
MYAVPQWIEYAVRVRGHTGFSFLYAKRASVILKRAEDFNIIAHIGNAGSCCV
KNGVSDTSMGITPLBGLIMTRCGDCDPALAFYINRKTGMSVSDITTLNKGSLIG
ITGKYDRDRCAAMKNGDALARLAFDMEVHRIKRYFGAYVAALQKQPDALVTAGVG
EMCCDVRACAGLEHLGRLDARKNELARNAETEISDDSPVRLIVPTDSELVN
TEDYALMOGTVDVHTRFTYSFQDPTTYNKAAGLKRDLKPNALATIVVPRLDT"
complement(4739..5464)
/gene="Tp0477"
/CDS
complement(4739..5464)
/gene="Tp0477"
/notes="similar to GP:1651209 percent identity: 37.04;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase, putative"
/protein_id="AAC65464.1"
/db_xref="GI:3322765"
/translation="MKKHIFEDARAIAAFLVSFDSRLKTOEILWLALSGSGSTPREIF
RWAHFRRHLDMKRLRFWSDERCVPPTDAQSNFMTHSALLEPLEINPAVPRVG
EAPASACAAYSQIEARLPRGVCPCFDIILLGADGHTASIFPHEIELWDHSGCC
VWATHPDTCQKRSVSTGHLINNAHEIYVVVTGREKQDMLASVASDPHNSVPLHAWTLP
KRSQGWIPLLRLGYPRKQKPTV"
complement(5525..7072)
/gene="Tp0478"
/CDS
complement(5525..7072)
/gene="Tp0478"
/notes="similar to GB:LA2023 SP:P44311 PID:1004000
PID:122494 PID:1204809 percent identity: 49.38;

```

```

identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase (zwf)"
/protein_id="AAC65465.1"
/db_xref="GI:3322766"
/translation="MGKISGSGTVAPHILVIFGASGLAARLKLIPLSLWDLFEQELLPR
TFGILGAGRTALSTESPRARAEAVTKHAYTDPHDPARLTFEQLOKHYSFSDTDSVA
FADPAAVYRTLDQSLHTEGNEFIVLATPPSLYETIPTQLAMHHLNREQGNFRVTEK
PEGYNLETAQHNLASLRAHFOENQTYRIDHYLGKETQONILVTFRANPLFEPTWNRTH
IDYVEITASISGVNRRGYDDQSGALRDMIONHLLLLGI IAMEAPAVYSSRLRDE
LVKVFDCLRPMGERDVMQHTVRAQYVAGKIRGVAVPGYLESGVDPSCPTETFAALKC
YIDNRMWDPVFIYRTGKRLPTGVTEIVHVRTLPALFHEIERPCAREGNALVIRIQ
PDEGTQLKIDLKEPGAGFKTIPVSVDFQYSALTYSHLPSAYERLLDCMNGDNWTLVHR
DDAVESAWRETDPIILAAWKSNNKSPLLTPAGSWGPKAADLINGSAPRWHPSSTILLS
DDEACRL"
complement(7145..7819)
/gene="Tp0479"
/CDS
complement(7145..7819)
/gene="Tp0479"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="r. pallidum predicted coding region TP0479"
/protein_id="AAC65468.1"
/db_xref="GI:3322769"
/translation="MTLSHPTRDRSRRLVQLGLDRMGIQWHITARGGVARCVVRVG
SLAALGMGAGVAAAEVFSVKVQVGAHPWKEPAPRTDVLITYTTLGLAVGVIAHN
GFAPTTDLDAFLMFAQLVGAIRGSKNFAFIPSLGVDILSTQDDKLVGIPITNL
DLIFFFTKHGLDFTFGSGVNVPLTRKNLKGANGQDLTWSINAWKCYCTSLVMVKI
GPVPR"
7931..8419
/gene="Tp0480"
/CDS
7931..8419
/gene="Tp0480"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="r. pallidum predicted coding region TP0480"
/protein_id="AAC65469.1"
/db_xref="GI:3322770"
/translation="MIRALFSLRSLHANTHPADLAHAALALALALLPRSSLLWYLL
FVCFIRLNRLGLLSLVLFGEVVPFSDPDLDSGNWALCLPRLQPVYRALIEIPFV
GLARFYNTIAGSLVAGALCYLPCYALARCAVTAVRTVLYPKIHHTATIFELVRNAPLC
KR"
8451..9884
/gene="Tp0481"
/CDS
8451..9884
/gene="Tp0481"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="r. pallidum predicted coding region TP0481"
/protein_id="AAC65470.1"
/db_xref="GI:3322771"
/translation="MSDDSTPKTPSRRIHTRRRRLHRRFFCKRVTPRSLKRLRRTH
IPADRAYCMRYLADPVSTPVVRVGTLLSRVVRDQAIASADLKRNLNIAASIAK
OGRVNWFLSNACASVLALGLVYLIRNIVARRVYIGSGEAFGAKCEAAVVDLDLF
NARFLKNVAVANKHHPWNLFIEISIDIHFDLLELSRGKEVSHVMVETWTPRK
TSGALPPRRKQRVRSSNPLIAKTOEKAAALAAAPVSGAGFSALKQAQDPFRLLERE
VKALQTLVQHVGAQAPKLAQTPQVDEAHARAETKAAIIRAVTELDJHAKLDVSA
IKQETLDRARRSTELATATISHLEQDQVHSTLGLAREFAAAVADGARTARAA
AAIRQIDQGGKKFISGLCTVFLARFSHYYPYVAQMLDYVRGSGQRTPSDGSFAEAE
KTAOSLTTRKRLOGVIFCLSATSLPC"
9851..10699
/gene="Tp0482"
/CDS
9851..10699
/gene="Tp0482"
/notes="hypothetical protein; identified by Glimmer;

```

Sat Nov 4 18:11:41 2000

```

Query Match      12.1%; Score 36.2; DB 1; Length 15727;
Best Local Similarity 51.6%; Pred. No. 2.2;
Matches 83; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 63 ctggcaactactcttgtaattgtcatttctaaagcggagggacacacctgaacccgcaat 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2650 CTGAGACACGCGCTTTTATTGTGTAFCNAAGAGAGTGTAGCAGCTTGAACACTCAGT 2591

QY 123 ggtctactagaagtacagaaagccttcagagatcggggctgcaattctcgaaacagggt 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2590 AATGAGCTTTTCCTCGCACACGCTCTTCTGTCAGCAGGTCTAGAACCGCACACACAGTTC 2531

QY 183 gttgcaattactcaagaaattctctgttggaataacactgc 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2530 GTGGCAGTTACCAGCGAGACCACTGCTCACTTGCATAATAATCC 2490

RESULT 8
AL365270/c      156995 bp      DNA      HTG      07-JUL-2000
LOCUS          Homo sapiens chromosome 1 clone RP11-99A8, *** SEQUENCING IN
DEFINITION     PROGRESS ***, 16 unordered pieces.
ACCESSION      AL365270
VERSION        AL365270.1 GI:8980271
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 156995)
AUTHORS        Pavlitt,R.
TITLE          Direct Submission
JOURNAL        Submitted (06-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
                CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
                requests: clonerequest@sanger.ac.uk
                ----- Genome Center
COMMENT        Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquery@sanger.ac.uk
                ----- Project Information
                Center project name: BA99A8
                ----- Summary Statistics
                Assembly program: XGAP4; version 4.5
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 18% of reads Chemistry:
                Dye-terminator Big Dye; 81% of reads
                Consensus quality: 148967 bases at least Q40
                Consensus quality: 152162 bases at least Q30
                Consensus quality: 154190 bases at least Q20
                Insert size: 155495; sum-of-contigs
                Insert size: 143559; agarose-fp
                Quality coverage: 4.30x in Q20 bases; sum-of-contigs Quality
                coverage: 4.65x in Q20 bases; agarose-fp
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 16 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                * 1 10300: contig of 10300 bp in length
                * 10301 10400: gap of 100 bp
                * 10401 13921: contig of 3521 bp in length
                * 13922 14021: gap of 100 bp
                * 14022 27982: contig of 13961 bp in length
                * 27983 28082: gap of 100 bp
                * 28083 30199: contig of 2117 bp in length
                * 30200 30299: gap of 100 bp
                * 30300 36658: contig of 6359 bp in length
                * 36659 36758: gap of 100 bp

```

## FEATURES

```

source
1..156995
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RPC1-11.1"
1..10300
/note="assembly_fragment:00909
fragment_chain:1"
10401..13921
/note="assembly_fragment:00428
fragment_chain:1"
14022..27982
/note="assembly_fragment:00957
fragment_chain:1"
28083..30199
/note="assembly_fragment:00712
fragment_chain:2"
30300..36658
/note="assembly_fragment:00682
fragment_chain:2"
36759..61477
/note="assembly_fragment:00988
fragment_chain:3"
61578..75491
/note="assembly_fragment:00212
fragment_chain:3"
75592..90911
/note="assembly_fragment:01372
fragment_chain:4"
91012..99070
/note="assembly_fragment:00923
fragment_chain:4"
99171..101293
/note="assembly_fragment:00031"
101394..109454
/note="assembly_fragment:00341"
109555..110702
/note="assembly_fragment:00691"
110803..112576
/note="assembly_fragment:00755"
112677..139381
/note="assembly_fragment:01073"
139482..148660
/note="assembly_fragment:01141"
148761..156995
/note="assembly_fragment:00305
clone_end:SP6
vector_side:right"
BASE COUNT 45192 a 31015 c 30671 g 48611 t 1506 others
ORIGIN

```

```

Query Match      11.9%; Score 35.6; DB 85; Length 156995;
Best Local Similarity 52.7%; Pred. NO. 4.5;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 120 aatggtctactagaatcacagaacccctcagagatcggggctgcaattctcgaacag 179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 46917 AAAGGTTTACAGGAGATGGAACGATATAGAGAAATCTGAAGGTCACTAACAAAG 46858
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 180 ggtgttgcaattactcaagaataattctctgttgataacactgctagaatagaggatgg 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 46857 GAAAGACCAATAGAAATTGTAATATCCAGGTAAATAAATAGAGTATTATTGCCCTCT 46798
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 240 ttatctcggtttcctctatgttgattg 265
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 46797 TAAGTTGTATATATGATGATTG 46772
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
AL353604/c
LOCUS      AL353604      101370 bp      DNA      HTG      09-AUG-2000
DEFINITION Homo sapiens chromosome 1 clone RP5-1180C18 map p32.3-34.2, ***
SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION      AL353604
VERSION      AL353604.9      GI:9795005
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Wall,M.
Direct Submission
Submitted (08-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 11, 2000 this sequence version replaced gi:9717057.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj1180C18
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ABI; 3% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Chemistry: Dye-terminator ET-amersham; 50% of reads
Quality: 99056 bases at least Q40
Consensus quality: 100022 bases at least Q30
Consensus quality: 100489 bases at least Q20
Insert size: 101270; sum-of-contigs
Insert size: 114210; 4.1% error; agarose-fp
Quality coverage: 0.00x in Q20 bases; sum-of-contigs Quality
coverage: 0.00x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 76800: contig of 76800 bp in length
* 76901 101370: contig of 24470 bp in length.
      Location/Qualifiers
        1..101370
          /organism="Homo sapiens"

/db_xref="taxon:9606"
/chromosome="1"
/map="p32.3-34.2"
/clone="RP5-1180C18"
/clone_lib="RPCI-5"
1..76800
/note="assembly_fragment:00905.0"
76901..101370
/note="assembly_fragment:01232.0"

BASE COUNT 24879 a 24116 c 25164 g 27111 t 100 others
ORIGIN

Query Match      11.8%; Score 35.4; DB 64; Length 101370;
Best Local Similarity 49.7%; Pred. No. 4.9;
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 120 aatggtctactagaatcacagaacccctcagagatcggggctgcaattctcgaacag 179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60814 AATGCGATGAAGGAGTCAAGGAAGATGCCACAGATGCTGCCCTGAGAAACTGAAAAAT 60755
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 180 ggtgttgcaattactcaagaataattctctgttgataacactgctagaatagaggatgg 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60754 GGAGCTTCGTAAACCAATAAATTTACCTTTTATGTTTAAAGCCCAAGTTGAAGTGGGTTTA 60695
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 240 ttatctcggtttcctctatgttgattggttggtggtgacttcagaaaatgtctgctgtg 299
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60694 TCACTTGGAAATTTTTTTTAAATATTGAGTAATTCAGAGACTACCAACAGCCAGAGGTA 60635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 g 300
Db 60634 G 60634

RESULT 10
AL353604/c
LOCUS      AL353604      11836 bp      DNA      BCT      26-MAY-2000
DEFINITION Chlamydia muridarum, section 83 of 85 of the complete genome.
ACCESSION      AE002356 AE002160
VERSION      AE002356.2      GI:8163342
KEYWORDS
SOURCE      Chlamydia muridarum.
ORGANISM      Chlamydia muridarum
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE      1 (bases 1 to 11836)
AUTHORS      Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
              White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
              Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
              Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
              McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
              Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
              pneumoniae AR39
              Nucleic Acids Res. 28 (6), 1397-1406 (2000)
REFERENCE      2 (bases 1 to 11836)
AUTHORS      Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
              White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
              Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
              Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
              McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
              Direct Submission
              Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA
              On Jun 1, 2000 this sequence version replaced gi:7190921.
              Location/Qualifiers
                1..11836
                  /db_xref="taxon:83560"
                  /note="synonym: Chlamydia trachomatis MoPn"
                  /complement(69..1301)
                  /gene="TC0894"
                  /complement(69..1301)
                  /gene="TC0894"

FEATURES
source
gene
CDS

```





```

/Note="assembly_fragment"
1674. .3543
/Note="assembly_fragment"
3644. .5601
/Note="assembly_fragment"
5702. .7732
/Note="assembly_fragment"
7833. .10667
/Note="assembly_fragment"
10768. .13917
/Note="assembly_fragment"
14018. .17381
/Note="assembly_fragment
clone_end:T7
vector_side:right"
17482. .20415
/Note="assembly_fragment"
20516. .23137
/Note="assembly_fragment"
23238. .27089
/Note="assembly_fragment
clone_end:SP6
vector_side:left"
27190. .33459
/Note="assembly_fragment"
33560. .39422
/Note="assembly_fragment"
39523. .46446
/Note="assembly_fragment"
46547. .56042
/Note="assembly_fragment"
56143. .67767
/Note="assembly_fragment"
67868. .86875
/Note="assembly_fragment"
86976. .106760
/Note="assembly_fragment"
106861. .132195
/Note="assembly_fragment"
132296. .162149
/Note="assembly_fragment"
50445 a 30230 c 30351 g 49322 t 1801 others
BASE COUNT 50445 a 30230 c 30351 g 49322 t 1801 others
ORIGIN

Query Match 11.7%; Score 35; DB 48; Length 162149;
Best Local Similarity 61.5%; Pred. No. 6.8;
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 207 ctgttgataacactgtagaataagaggatggttagctcggttctctatgtttgattgg 266
Db 116844 CTGTTAAATTAATAATGCTACACATCAGTCTGTTTACCTGTTTATTTATTTTCATTCGG 116785

Qy 267 gttgtgttaggacttcagaaatgtctgtcgt 297
Db 116784 ATTTGGGTATGACTATACATGATCTGCTG 116754

RESULT 12
AC019951/c 33328 bp DNA HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
DEFINITION pieces
ACCESSION AC019951
VERSION AC019951.1 GI:6664946
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 33328)
AUTHORS Adams,M. and Venter,J.C.

Direct Submission
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10211508 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. .33328
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 9249 a 7305 c 7379 g 9395 t
ORIGIN

Query Match 11.5%; Score 34.6; DB 40; Length 33328;
Best Local Similarity 49.2%; Pred. No. 7.6;
Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 94 agagcggaggcacacacgtgaaccccgcaatggtctactaagatcacagaagccttcagag 153
Db 12366 ATAGCATTTGAATCATTCACACCTGCTATTATTTAGCCGGAATTAACAAGACTCTTCTCTG 12307

Qy 154 atcggggctgcaattctgaacagggtgttgcattactcaagaaaatctctgttgg 213
Db 12306 ATACGGATATATGCTATACATTTGAGGCTTTTATATTAAGATAAATTTTAAATAA 12247

Qy 214 ataacactgctagaatagaggatggttagctcggttctctatgtttgattgggttg 273
Db 12246 GTCTGCTTACAAAATAGATCCCAATTCATCTGGGTTTTTTTAACTCTGCTAGTCGCG 12187

Qy 274 gtagg 278
Db 12186 CGAGG 12182

RESULT 13
AC010575 148447 bp DNA HTG 18-FEB-2000
LOCUS Drosophila melanogaster clone RPC198-5W7, *** SEQUENCING IN
DEFINITION PROGRESS ***, 52 unordered pieces.
ACCESSION AC010575
VERSION AC010575.3 GI:6996792
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 148447)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichter,A., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu.L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Suchang,R.,
Tabor,P., Taylor,T., Vasquez,I., Vinson,R., Vo,Q., Wahbah,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished

```



## REFERENCE

2 (bases 1 to 148447)

Worley, K.C.

Direct Submission

Submitted (16-SEP-1999)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Feb 18, 2000 this sequence version replaced gi:5902992.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: DRGW

Center clone name: RCI98-5M7

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 17% of reads

Assembly program: Phrap; version 0.980611

Consensus quality: 105301 bases at least Q40

Consensus quality: 119508 bases at least Q30

Consensus quality: 126701 bases at least Q20

Estimated insert size: 132501; sum-of-coverage estimation

Quality coverage: 0.6x in Q20 bases; sum-of-coverage estimation

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 52 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 850: contig of 850 bp in length  
 \* 851: gap of unknown length  
 \* 871: 1717: contig of 847 bp in length  
 \* 1718: 1737: gap of unknown length  
 \* 1738: 2614: contig of 877 bp in length  
 \* 2615: 2634: gap of unknown length  
 \* 2635: 3454: contig of 820 bp in length  
 \* 3455: 3474: gap of unknown length  
 \* 3475: 4393: contig of 919 bp in length  
 \* 4394: 4413: gap of unknown length  
 \* 4414: 5289: contig of 876 bp in length  
 \* 5290: 5309: gap of unknown length  
 \* 5310: 6141: contig of 832 bp in length  
 \* 6142: 6161: gap of unknown length  
 \* 6162: 6984: contig of 823 bp in length  
 \* 6985: 7004: gap of unknown length  
 \* 7005: 8080: contig of 1076 bp in length  
 \* 8081: 9327: contig of 1227 bp in length  
 \* 9328: 9347: gap of unknown length  
 \* 9348: 10228: contig of 881 bp in length  
 \* 10229: 10249: gap of unknown length  
 \* 10249: 11302: contig of 1054 bp in length  
 \* 11303: 11322: gap of unknown length  
 \* 11323: 12174: contig of 852 bp in length  
 \* 12175: 13297: contig of 1103 bp in length  
 \* 13298: 13317: gap of unknown length  
 \* 13318: 14539: contig of 1222 bp in length  
 \* 14540: 14559: gap of unknown length  
 \* 14560: 15698: contig of 1139 bp in length  
 \* 15699: 15718: gap of unknown length  
 \* 15719: 16522: contig of 804 bp in length  
 \* 16523: 16542: gap of unknown length  
 \* 16543: 17518: contig of 976 bp in length  
 \* 17519: 17538: gap of unknown length  
 \* 17539: 18404: contig of 866 bp in length  
 \* 18405: 18424: gap of unknown length  
 \* 18425: 19389: contig of 965 bp in length  
 \* 19390: 19409: gap of unknown length

19410 20870: contig of 1461 bp in length  
 \* 20871: gap of unknown length  
 \* 20891: 22733: contig of 1845 bp in length  
 \* 22736: 22755: gap of unknown length  
 \* 22756: 24425: contig of 1670 bp in length  
 \* 24426: 24445: gap of unknown length  
 \* 24446: 25358: contig of 913 bp in length  
 \* 25359: 25378: gap of unknown length  
 \* 25379: 26724: contig of 1346 bp in length  
 \* 26725: 26744: gap of unknown length  
 \* 26745: 28410: contig of 1666 bp in length  
 \* 28411: 28430: gap of unknown length  
 \* 28431: 29885: contig of 1455 bp in length  
 \* 29886: 29905: gap of unknown length  
 \* 29906: 30781: contig of 876 bp in length  
 \* 30782: 30801: gap of unknown length  
 \* 30802: 31715: contig of 914 bp in length  
 \* 31716: 31735: gap of unknown length  
 \* 31736: 32774: contig of 1039 bp in length  
 \* 32775: 32794: gap of unknown length  
 \* 32795: 34403: contig of 1609 bp in length  
 \* 34404: 34423: gap of unknown length  
 \* 34424: 35892: contig of 1469 bp in length  
 \* 35893: 35912: gap of unknown length  
 \* 35913: 37899: contig of 1987 bp in length  
 \* 37900: 37919: gap of unknown length  
 \* 37920: 39450: contig of 1531 bp in length  
 \* 39451: 39470: gap of unknown length  
 \* 39471: 41240: contig of 1770 bp in length  
 \* 41241: 41260: gap of unknown length  
 \* 41261: 43877: contig of 2617 bp in length  
 \* 43878: 43897: gap of unknown length  
 \* 43898: 45845: contig of 1948 bp in length  
 \* 45846: 45865: gap of unknown length  
 \* 45866: 48533: contig of 2668 bp in length  
 \* 48534: 48553: gap of unknown length  
 \* 48554: 50809: contig of 2256 bp in length  
 \* 50810: 50829: gap of unknown length  
 \* 50830: 55053: contig of 4224 bp in length  
 \* 55054: 55073: gap of unknown length  
 \* 55074: 58837: contig of 3764 bp in length  
 \* 58838: 58857: gap of unknown length  
 \* 58858: 62191: contig of 3334 bp in length  
 \* 62192: 62211: gap of unknown length  
 \* 62212: 65883: contig of 3672 bp in length  
 \* 65884: 65903: gap of unknown length  
 \* 65904: 70458: contig of 4555 bp in length  
 \* 70459: 70478: gap of unknown length  
 \* 70479: 75989: contig of 5511 bp in length  
 \* 75990: 76009: gap of unknown length  
 \* 76010: 82490: contig of 6481 bp in length  
 \* 82491: 82510: gap of unknown length  
 \* 82511: 88156: contig of 5646 bp in length  
 \* 88157: 88176: gap of unknown length  
 \* 88177: 94440: contig of 6264 bp in length  
 \* 94441: 94460: gap of unknown length  
 \* 94461: 100652: contig of 6192 bp in length  
 \* 100653: 100678: gap of unknown length  
 \* 100679: 110072: contig of 9406 bp in length  
 \* 110079: 110098: gap of unknown length  
 \* 110099: 121746: contig of 11648 bp in length  
 \* 121747: 121766: gap of unknown length  
 \* 121767: 148447: contig of 26681 bp in length.

## FEATURES

source

Location/Qualifiers

1. 148447

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="RCI98-5M7"

42418 a 32481 c 31053 g 41428 t 1067 others

BASE COUNT

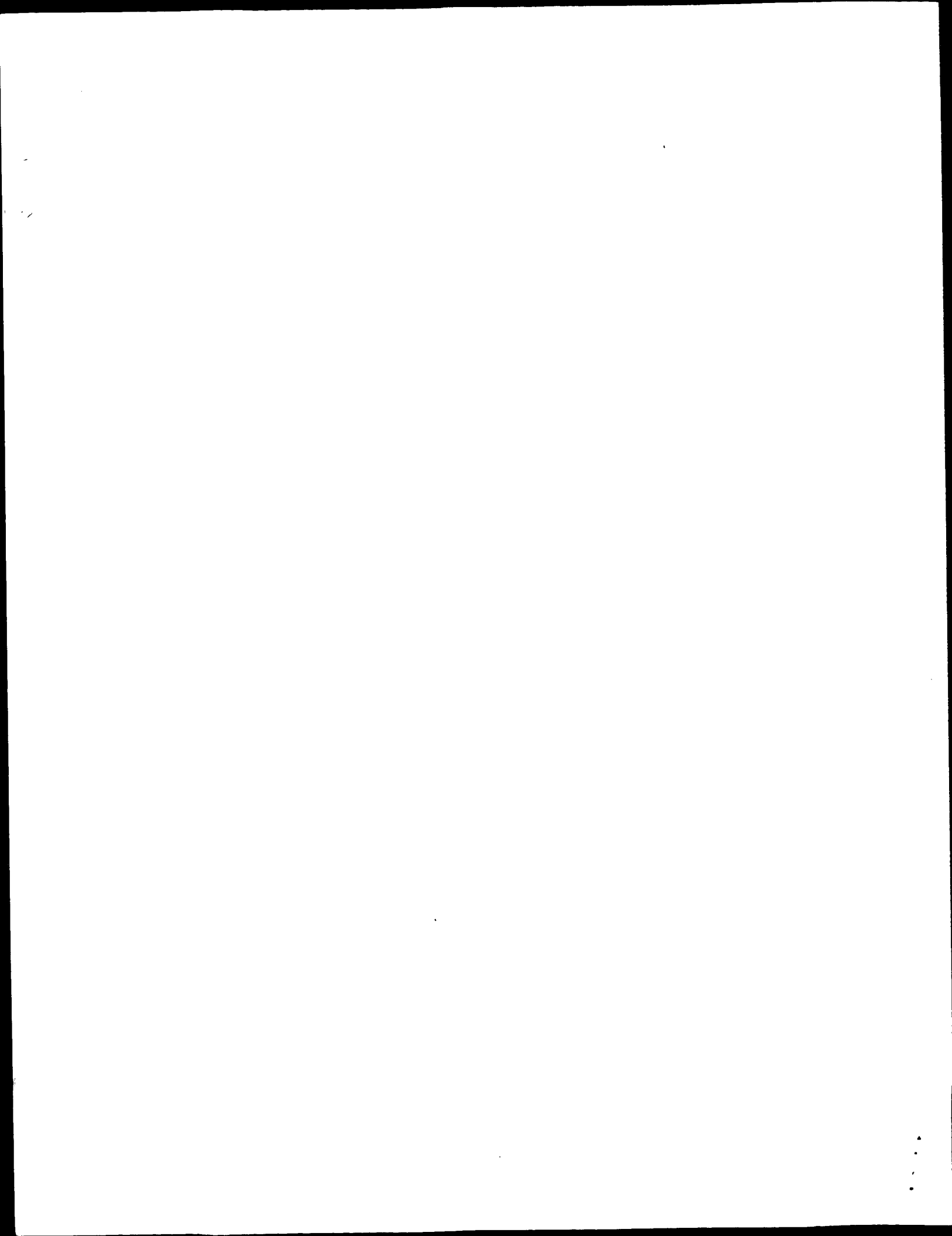
ORIGIN

Query Match

11.5%; Score 34.6; DB 36; Length 148447;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:32:00 ; Search time 4075.18 Seconds  
(without alignments)  
245.431 Million cell updates/sec

Title: US-09-300-482-27

Perfect score: 229

Sequence: 1 cagacctatttttttctgc.....cttcaatttatggggaacaa 229

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_om:\*

4: gb\_ov:\*

5: gb\_pat:\*

6: gb\_ph:\*

7: gb\_pll:\*

8: gb\_pl2:\*

9: gb\_pr1:\*

10: gb\_pr2:\*

11: gb\_pr3:\*

12: em\_fun:\*

13: em\_hum1:\*

14: em\_hum2:\*

15: em\_in:\*

16: em\_om:\*

17: em\_or:\*

18: em\_ov:\*

19: em\_pat:\*

20: em\_ph:\*

21: em\_pl:\*

22: em\_ro:\*

23: em\_sts:\*

24: em\_sy:\*

25: em\_un:\*

26: em\_vi:\*

27: gb\_htg1:\*

28: gb\_htg2:\*

29: gb\_in1:\*

30: gb\_in2:\*

31: em\_ba1:\*

32: em\_ba2:\*

33: em\_hum3:\*

34: em\_hum4:\*

35: gb\_pr4:\*

36: gb\_htg3:\*

37: gb\_htg4:\*

38: gb\_htg5:\*

39: gb\_htg6:\*

40: gb\_htg7:\*

41: em\_htg1:\*

42: em\_htg2:\*

43: em\_htg3:\*

44: em\_hum5:\*

45: gb\_pl3:\*

46: gb\_pr5:\*

47: gb\_htg8:\*

48: gb\_htg9:\*

49: gb\_htg10:\*

50: gb\_htg11:\*

51: gb\_htg12:\*

52: gb\_htg13:\*

53: gb\_htg14:\*

54: gb\_in3:\*

55: gb\_htg15:\*

56: gb\_htg16:\*

57: gb\_htg17:\*

58: em\_htg4:\*

59: em\_htg5:\*

60: em\_htg6:\*

61: em\_htg7:\*

62: em\_hum6:\*

63: gb\_htg18:\*

64: gb\_htg19:\*

65: gb\_ba3:\*

66: em\_htg8:\*

67: em\_htg9:\*

68: em\_htg10:\*

69: em\_htg11:\*

70: em\_htg12:\*

71: em\_htg13:\*

72: em\_htg14:\*

73: em\_htg15:\*

74: em\_htg16:\*

75: em\_htg17:\*

76: em\_htg18:\*

77: em\_htg19:\*

78: em\_htg20:\*

79: em\_htg21:\*

80: em\_htg22:\*

81: em\_htg23:\*

82: gb\_pr6:\*

83: gb\_pr7:\*

84: gb\_htg20:\*

85: gb\_htg21:\*

86: gb\_htg22:\*

87: gb\_htg23:\*

88: gb\_ro:\*

89: gb\_sts1:\*

90: gb\_sts2:\*

91: gb\_sy:\*

92: gb\_un:\*

93: gb\_vil:\*

94: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	136.2	59.5	1667	7	AB007907 Glycine m
2	84.2	36.8	1618	45	U18239 Medicago sa
3	77	33.6	96240	56	AC068900 Arabidops
4	60.2	26.3	634	7	AF037030 Zea mays
5	60.2	26.3	150120	8	AF001552 Oryza sat
6	58.6	25.6	1686	7	AF061838 Zea mays
7	57.4	25.1	1818	7	AF061837 Zea mays
8	57.4	25.1	2392	7	AF037037 Zea mays
9	55.8	24.4	645	7	AF037029 Zea mays
10	52.4	22.9	111222	7	AC007764 Genomic s
11	50.8	22.2	79837	7	AB005233 Arabidops
12	38.2	16.7	170453	10	AC006600 Homo sapi





SOURCE	ORGANISM	FEATURES	source
Zea mays.	Zea mays		
Eukaryota;	Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.		
REFERENCE			
AUTHORS	Padegimas, L.S. and Reichert, N.A.		
TITLE	Isolation of genes and regulatory sequences implicated in hypersensitive response from Zea mays nematode-resistant line MP307 unpublished		
JOURNAL	2 (bases 1 to 634)		
REFERENCE			
AUTHORS	Padegimas, L.S. and Reichert, N.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-DEC-1997) Department of Plant and Soil Sciences, Mississippi State University, Box 9555, Mississippi State, MS 39762, USA		
FEATURES	Location/Qualifiers		
source	1..634		
CDS	/organism="Zea mays"		
	/db_xref="taxon:4577"		
	100...>634		
	/note="8pgdh(B)"		
	/codon_start=1		
	/product="6-phosphogluconate dehydrogenase isoenzyme B"		
	/protein_id="AAC79950.1"		
	/db_xref="GI:3925225"		
	/translation="MALTRIGIAGLVAMQNLNIAEKFPFISVYNFTTSKVETVQ RAKRGNLPVTGFHPDPAFVSNIQKPRVVIMLVKAGAPVDQTATLAAHLEQGDCIID GNGERYENTERREKKAMEERGLLYLGMGVSGGEGARNGPSLMGGSPDAYKYVEDIVL KVAAGVPDSDGPCVTYIGR"		
BASE COUNT	139 a 174 c 204 g 117 t		
ORIGIN			
Query Match	26.3%; Score 60.2; DB 7; Length 634;		
Best Local Similarity	67.1%; Pred. No. 1.6e-07;		
Matches	100; Conservative 0; Mismatches 48; Indels 1; Gaps 1;		
Qy	65 acagaataagccttgcgtgattggctgttaatggcgcaaaatctggcacactaatattgct 124 		
Db	109 ACAAGATCGTCTTGTCGCCCTCGCTGTGATGGGGCAGAAACCTTGCCCTCAACATTGCA 168		
Qy	125 tgaaaagggtttcccaattccggtttaacaacgaagaacatttcgaaggttatgggccata 184 		
Db	169 GAGAAAGGG-TTCCCATCTCTGTGTACACAGACAACTCCAAGGTAGACGAGACC GT 227		
Qy	185 agacgaagcaaacaccaggaaggaaccttc 213 		
Db	228 GCACGTGCCAAGCAGCAGGAACCTTC 256		
RESULT 5			
AP001552	LOCUS AP001552 150120 bp DNA PLN 16-JUN-2000		
DEFINITION	Oryza sativa genomic DNA, chromosome 6, PAC clone:P0029D06.		
ACCESSION	AF001552		
VERSION	AP001552.1 GI:7363267		
KEYWORDS	.		
SOURCE	Oryza sativa (cultivar:Nipponbare) DNA, clone:P0029D06. Oryza sativa		
ORGANISM	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.		
REFERENCE	1 (bases 1 to 150120)		
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.		
TITLE	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC clone:P0029D06		
JOURNAL	Published Only in DataBase (2000) In press		
REFERENCE	2 (bases 1 to 150120)		
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@agr.affrc.go.jp,		



DGDRRGGDSRRRRHIAATMRGAARARCEGEGVLTGVRPATRETTNGDAA  
 BLOTLREEMREVRGREGVGEREPTTAAMGLTGTTRTTFADGGIRWRGRRAAA  
 SGRRSCKGGGRNDRGHGRWRKEEGEGEAL"  
 join(19096...19172,20123...20329,20389...20395)  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAA93017.1"  
 /db\_xref="GI:7363273"  
 /translation="MOLEPTMHQCLVOWGDKLIEVPANRPOSCSRRRLGTAAPTRRP  
 DIPLPSPESRRRTLPAAANAAPSRPPASGQRHSISGSGSALRLSGLRI"  
 21587...21769  
 /note="5' LTR"  
 join(21843...22489,22572...23072,23168...24293)  
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,  
 BAC clone F17A8; putative protein. (AL049482)"  
 /codon\_start=1  
 /protein\_id="BAA93018.1"  
 /db\_xref="GI:7363274"  
 /translation="MEKSIALLFOKHEAKKNACDLPIPHVDEKSGSIPAVNDAPIDED  
 IIVDEGVETEEDTVDDPPVDDVDEVTETKEESLPYVDVLDLHDPGLVPLIS  
 SFANDQDAARRGYILKPGCHLWAFNFPKRYGKDRFSTVWEHKYPIWISYVDKDA  
 TFCFVCLFEGESGKFTVGTGHNWANNVAKALDKHVGCTSSDHNFAQRYNLFVKKGL  
 RYLLROGLAFRGHDETESNRCNFELELLKWLAGNENANKVVLNNAFGNCLTSPRI  
 OKQVIRCCSWETSCILKELGDEHYAILADESSDISHKEQLWCLRYVDKLGVCERF  
 LAVVAGTSLLQATAIQSLTSHLLTQIHGGQYDGSNMKGVEYKGCWLPNLFQ  
 QVRLNIIVGVSCTRDHMDLVRAQKLKALNLGIESGSLNGOEMALARPGLDRWGS  
 HYKTLIHIDMTDITREVLITIGKPTOREDWPIIHAMVLAFESFEVFNALHMLYL  
 GTFNFSNLSQKRDODIVNMSLVGLAKKMOEFENGWEGFLKVSFCIKYSIDIP  
 AMDKVPVHGHRSHRFYPTQIDHVRERYIGVIRHQELENRFDESMELLCSA  
 FNPTDSFASDAQILRLASFYKPDIEGNLMLQELQDITYINDREDHRFKGLNKIG  
 QLSIKLVEFKKHDLDVLLKLLILFVATASVERFVSAMNLVKTVRNMSMSDKLL  
 NCLVTFIERDMYMRKT"  
 26421...26603  
 /note="3' LTR"  
 complement(join(29217...30334,31321...31609,31701...31862))  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAA93019.1"  
 /db\_xref="GI:7363275"  
 /translation="MYQIKIRKKRDEVLVATRLGKVESIALRLMDGRVSGDDAGE  
 RRLPRRAACGPKSKVLGDLWGLPWITSLTHPPRLASTPIGRAAAANAATAGL  
 VSSPAHCKGLRVKRWGAGSGEDHGRRRRPRSPRSPSLKNSHMLHQI  
 IALLSSFDLLAFSGTCSWRAALSSPFIYTFPPLHLKPDIPNSHPHCSFRYTL  
 LYKPKWGLDSKRLSUCAPONTNRMRYLGSYGLIFSYENCLLDVMTAK  
 VPKPLQAGAKETTYGILAPLNLPIHLLCSRSSIFYWQVTSWSHPLVFGRI  
 LQILVSGEFTAMFPHRLHTMREAPQLSMQEVGVWGEEMFVGHFKWLVISGDM  
 LMLDSVGIHHSYGPFGTFQVRLDFSAQAKWMKLENSALFVSLDRNPFTSCT  
 NPERWKGKNCIYVAKPSEDSDPWTAVELGQPIPGATHCVPSHPLLRTEGHCQLE  
 YLWLPSPINGVDQ"  
 32671...33849  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAA93020.1"  
 /db\_xref="GI:7363276"  
 /translation="MQSPKVSSEAGSPATSVIEGWAEPLPCLLSHISVALLGSLFD  
 LLAFTGTCHSWRAAFSSYPKSNPTILPLLVRPNVRVKAIPSSHGRKLRSCEVID  
 LANRTPLCQIPQETLQRHFGAGSGGLICCRGYCLVDVFTGAESVPLRPFSE  
 NCDEFYCGILTAPITPSNHLIISTDSDWPSGSDSELKLPVNRVDQIVFENG  
 QLIAVIETSLTQALPIKLEIKTLWMDNNECPMRPFVVCVCDMLLIYDVHISF  
 SFGAVIYRPLRDMSTKPAKWEVKLENLWALFIGDARSPPSPFKPFRWGSRNC  
 LYAHYSQPLSLHGLGDADAVDNPNTDNLVFRNRYQLQALWVYPSMFSAGDQ  
 complement(join(34717...34935,35401...35912,36242...36363,  
 36470...36542,36626...36754,36936...37247,37336...38617,  
 38724...39212))  
 /note="Similar to Arabidopsis thaliana DNA chromosome V,  
 BAC clone TM021B04; N. tabacum membrane-associated  
 salt-inducible protein. (AF00721)"  
 /codon\_start=1  
 /protein\_id="BAA93021.1"  
 /db\_xref="GI:7363277"  
 /translation="MAAATAGAAASTAVSITGSSSYEDDGAATSWLSSTSSSSSS  
 SQRQPYRLLHDEAQLRRRERGQSGAHTPRVRRTPDQMARVYEDDRAGHYVGRH

VVAARAARATASCSPSSADMBEAMASFVAKLTIFREMFVLIHURGWQRADFFAWM  
 KLOLCVPSVAVTILLRLQGVKVLAEVTFLEMLAGCEPDVAAGCTLLCAVWR  
 GKNDMLMFAAARRDDIVPSIVFNFVSSLOKOLHGVKVIHLMEQLEANVAPNQF  
 TTVVIGSAGKMGLEAMDAFGEMRRFVPEATYSLILSLCAKHGKGEALGLYD  
 EMVKSVISNTYTCASVLTLYKNEDYSKALSLFSEMEONKIVPDEVYIGLIRLYGK  
 LGLYEADQRMEEDIDKAGLSDEQTVMAAQVHMVQNYDRLAOVLDMARARNVPSQ  
 FSYSALLRHCHVAKEDVDAEDTFRLSNYGPDPVCCNDLLRYLWRLHLDKARALIL  
 MKLEALQFQEDLCVTVLEVCCKTSINKDITDNLTEVIONEGSSKVLNPTDSSLMM  
 KSLLDKPGGLSSVLSOLIMKFAREGSTDEAKELYEHLTEAGKPDPTATILVIOGV  
 AQOLEAQKLFETASTFPVGGSVNAMYDALCRGKTEEAYRFLMEILDGHNDAV  
 TISILVHLTKQKFOEANIIRCLHDEALDTVYNTFIKMSLEGLDNVLLAVS  
 IYDRMLSSGIPRSMOTFIMISVYGQGLKXAVEMFAAOELGPIDEKTYTNMISF  
 YKAGHLHNEAIFQEMKNHVPDSHTYLALIRAVTEGKCYKAEATOMLNSMT  
 PSTCHNLISAFLEKQIDEAQRMYNOMEAGIPADLACCTWMAHLDHGYVDDGI  
 LFFETARLLKPDSTLSAAPHLYEHSGRESAGDYLDAINMSGFRLNLKVSCKLE  
 QIQPKWSIVKSGPLFPQKAHGVPLKYKORSVCCSEQVWAREGEVDAARFQSRDA  
 AVHDHVQRVDSSTCD"  
 complement(join(40023...40413,40474...40971,41293...41464,  
 41783...42027,42111...42288,42719...42901,43197...43301,  
 44137...44274,44361...45353))  
 /note="ESTs C74776(E51022),C26123(C116681) correspond to a  
 region of the predicted gene.  
 Similar to Arabidopsis thaliana cultivar Landsberg  
 extra-large G-protein (AF060942)"  
 /codon\_start=1  
 /protein\_id="BAA93022.1"  
 /db\_xref="GI:7363278"  
 /translation="NAGAGVGSNWEEMVRMLPPGTTFEAPANLDYSIALYDGP  
 PVSVELPRDVPDLPAITAPQVSGPLVFGNSGVVAVPVVRPMPVHRKQDAHRAE  
 PPVAAQGRRRSSSDVSAFQNEGVSDDSCSVQSESAHFHGGGRTAAQEGRR  
 ADVTVGTVEDSRYSKEDVDVSEYVAVTKKRGRTSCRCGRKWKESKESCIVCDA  
 RECSYCVLRAMSGPEGRKCTICIGQPIDESKRLGKGRSLRLLSPLEVRQILKA  
 EKECAQNLRLPEQLIINGFLNPDEMALLSCQRPQKLPGRVWYDKESLWKGEKE

Query Match 26.3%; Score 60.2; DB 8; Length 150120;  
 Best Local Similarity 64.2%; Pred. No. 1.8e-07;  
 Matches 106; Conservative 0; Mismatches 58; Indels 1; Gaps 1;  
 QY 65 acaagaatgagcctgctgattgctgtaataatggcgcaaaacttgcactcaattgct 124  
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50458 ACTAGAAATGGTCTGCTGGCTTCGGGTATGGGCGACACCTTGCCTCAACATGCA 50517  
 QY 125 tgaagaagcctcccaatcccggttaacacagcaaccatttcgaagttattggccata 184  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50518 GAGAAAGG-TTCCCTATCTGTCTACACAGGACGACTTCTAAGTGTGATGAGACGT 50576  
 QY 185 agacgaagcaaacaccaggaagaaaccttcaatttatggggaacaa 229  
 || || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 50577 TCAGCGCGCAAGGTAGAGGAACCTTCTCTGTACGGGTTCGA 50621  
 || || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 RESULT 6  
 AF061838 1686 bp mRNA PLN 08-JUN-2000  
 LOCUS zea mays strain W64A x A182E putative cytosolic 6-phosphogluconate  
 DEFINITION dehydrogenase (pdh2) mRNA, complete cds.  
 ACCESSION AF061838  
 VERSION AF061838.1 GI:3342801  
 KEYWORDS  
 SOURCE zea mays.  
 ORGANISM zea mays.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 1 (bases 1 to 1686)  
 Redinbaugh, M.G. and Campbell, W.H.  
 Nitrate regulation of the oxidative pentose phosphate pathway in  
 maize (Zea mays L.) root plastids: induction of 6-phosphogluconate  
 dehydrogenase activity, protein and transcript levels  
 Plant Sci. 134 (2), 129-140 (1998)  
 2 (bases 1 to 1686)  
 Redinbaugh, M.G. and Bailey-Serres, J.  
 Characterization of cDNAs encoding cytosolic and a putative  
 plastidic 6-phosphogluconate dehydrogenase



## JOURNAL

Submitted (06-DEC-1997) Department of Plant and Soil Sciences,  
Mississippi State University, Box 9555, Mississippi State, MS  
39762, USA

FEATURES  
source

Location/Qualifiers  
1. .2392  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
1858. .>2392  
/note="6pgdh(A)"  
/codon\_start=1  
/product="6-phosphogluconate dehydrogenase isoenzyme A"  
/protein\_id="AAC79957.1"  
/db\_xref="GI:3925239"  
/translation="MALTRIGLAGLAVMGONLALNIAEKGFPIISVYRTTSKVDETVQ  
RKVEGNLPVGFPHDPASVSSIQKPRVIMLVKAGAPVDQTIALAHLDDQGDICVD  
GGNEWENTERREKAMEERGLLYLGMVSGGEGARGNPSLMPGGSFEAYKYIEDILL  
KVAQVPSGPGCVTYIGK"

## CDS

642 a 529 c 546 g 668 t 7 others  
ORIGIN

Query Match 25.1%; Score 57.4; DB 7; Length 2392;  
Best Local Similarity 64.2%; Pred. No. 1.1e-06;  
Matches 102; Conservative 0; Mismatches 56; Indels 1; Gaps 1;  
QY 65 acaagaatagccttgcgtgattggtgttaattggcgaatactggcactcaattatgct 124  
Db 1867 ACAGAAATGGCTCGCGGCTCGCGGTCTACGACAGAACCTTCCCTCAACATCGC- 1925  
QY 125 tgaagaaggcttcccaattccggttaacaacggaacattcccaaggattatgggccata 184  
Db 1926 GGAGAAAGGTTCCCTCTACACAGGACAACTCCCAAGGTTGATGAGACCGT 1985  
QY 185 agacgaagcaaacaggaagaaacattcaattatggg 223  
Db 1986 GCACGTCGCAAGTCGAGGAAACCTCCACAGTGTGG 2024

## RESULT 9

AF037029 645 bp mRNA PLN 26-NOV-1998  
LOCUS  
DEFINITION Zea mays 6-phosphogluconate dehydrogenase isoenzyme A mRNA, partial cds.  
ACCESSION AF037029  
VERSION AF037029.1 GI:3925222  
KEYWORDS Zea mays.  
SOURCE Zea mays.  
ORGANISM Zea mays.

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.

Padegimas,L.S. and Reichert,N.A.  
Isolation of genes and regulatory sequences implicated in  
hypersensitive response from Zea mays line MP307

Unpublished  
2 (bases 1 to 645)  
Submitted (06-DEC-1997) Department of Plant and Soil Sciences,  
Mississippi State University, Box 9555, Mississippi State, MS  
39762, USA

FEATURES  
source

Location/Qualifiers  
1. .645  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
111. .>645  
/note="6pgdh(A)"  
/codon\_start=1  
/product="6-phosphogluconate dehydrogenase isoenzyme A"  
/protein\_id="AAC79949.1"  
/db\_xref="GI:392523"  
/translation="MALTRIGLAGLAVMGONLALNIAEKGFPIISVYRTTSKVDETVQ  
RKVEGNLPVGFPHDPASVSSIQKPRVIMLVKAGAPVDQTIALAHLDDQGDICVD

## CDS

642 a 529 c 546 g 668 t 7 others  
ORIGIN

Query Match 24.4%; Score 55.8; DB 7; Length 645;  
Best Local Similarity 63.5%; Pred. No. 3.3e-06;  
Matches 101; Conservative 0; Mismatches 57; Indels 1; Gaps 1;  
QY 65 acaagaatagccttgcgtgattggtgttaattggcgaatactggcactcaattatgct 124  
Db 120 ACAGAAATGGCTCGCGGCTCGCGGTCTACGACAGAACCTTCCCTCAACATCGC- 178  
QY 125 tgaagaaggcttcccaattccggttaacaacggaacattcccaaggattatgggccata 184  
Db 179 GGAGAAAGGTTCCCTCTACGACAGGACAACTCCCAAGGTTGATGAGACCGT 238  
QY 185 agacgaagcaaacaggaagaaacattcaattatggg 223  
Db 239 GCAGCGCGCAAGTCGAGGAAACCTCCACAGTGTGG 277

AC007764 111222 bp DNA PLN 28-JUN-2000  
LOCUS  
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F22C12 from chromosome I, complete sequence.  
ACCESSION AC007764  
VERSION AC007764.2 GI:5508844  
KEYWORDS HTG.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons, core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 111222)  
Shinn,P., Khan,S., Brooks,S., Buehler,E., Chao,Q., Dunn,P., Kim,C.,  
Walker,M., Altafi,H., Araujo,R., Conn,L., Conway,A.B., Gonzalez,A.,  
Hansen,N.F., Huizar,L., Kremenetskaia,I., Lenz,C., Li,J., Liu,S.,  
Luros,S., Rowley,D., Schwartz,J., Toriumi,M., Vysotskaia,V., Yu,G.,  
Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.  
Genomic sequence for Arabidopsis thaliana BAC F22C12 from chromosome I

Unpublished  
2 (bases 1 to 111222)  
Ecker,J.R.  
Direct Submission

Submitted (09-JUN-1999) Arabidopsis thaliana Genome Center,  
Department of Biology, University of Pennsylvania, 38th Street and  
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA

3 (bases 1 to 111222)  
Ecker,J.R.  
Direct Submission

Submitted (16-JUL-1999) Arabidopsis thaliana Genome Center,  
Department of Biology, University of Pennsylvania, 38th Street and  
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA

4 (bases 1 to 111222)  
Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,  
Shinn,P., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L.,  
Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,  
Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,  
Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,  
Thavari,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,  
Federspiel,N., Theologis,A. and Ecker,J.R.  
Direct Submission

Submitted (12-JAN-2000) Arabidopsis thaliana Genome Center,  
Department of Biology, University of Pennsylvania, 38th and  
Hamilton Walk, Philadelphia, PA 19104-6018, USA

5 (bases 1 to 111222)  
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,  
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,  
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,

Shinn,P., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L.,  
Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,  
Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,  
Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,  
Thavari,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,  
Federspiel,N., Theologis,A. and Ecker,J.R.  
Direct Submission

Submitted (12-JAN-2000) Arabidopsis thaliana Genome Center,  
Department of Biology, University of Pennsylvania, 38th and  
Hamilton Walk, Philadelphia, PA 19104-6018, USA

5 (bases 1 to 111222)  
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,  
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,  
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,

Shinn,P., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L.,  
Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,  
Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,  
Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,  
Thavari,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,  
Federspiel,N., Theologis,A. and Ecker,J.R.  
Direct Submission

Submitted (12-JAN-2000) Arabidopsis thaliana Genome Center,  
Department of Biology, University of Pennsylvania, 38th and  
Hamilton Walk, Philadelphia, PA 19104-6018, USA

5 (bases 1 to 111222)  
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,  
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,  
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,

Shinn,P., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L.,  
Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,  
Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,  
Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,  
Thavari,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,  
Federspiel,N., Theologis,A. and Ecker,J.R.  
Direct Submission

Submitted (12-JAN-2000) Arabidopsis thaliana Genome Center,  
Department of Biology, University of Pennsylvania, 38th and  
Hamilton Walk, Philadelphia, PA 19104-6018, USA

5 (bases 1 to 111222)  
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,  
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,  
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,

Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J.,  
Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H.,  
Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M.,  
Yu, G., Davis, R., Federspiel, N., Theologis, A., and Ecker, J.  
Direct Submission  
Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,  
Department of Biology, University of Pennsylvania, 38th and  
Hamilton Walk, Philadelphia, PA 19104-6018, USA  
On Jul 16, 1999 this sequence version replaced gi:5030434.

## FEATURES

Location/Qualifiers  
1..111222  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="F22C12"  
complement(join(979..3122,3735..5930,6862..6892,  
6983..7068,7655..9286,9346..10019,10228..10279,  
10673..10816,11047..13429,13585..13835,14366..14645))  
/note="similar to maize transposon MuDR mudra protein  
isolog gi|1946355"

## source

/codon\_start=1  
/evidence=not\_experimental  
/product="F22C12.1"  
/protein\_id="AAF24584.1"  
/db\_xref="GI:6692119"

## CDS

/translation="MKTGGLITFTVILLASLPPTKATGAGSRKRDRECEESRI  
QTCDVNVSLKISTECCKPLKQOCLCDVTKTKITNVLSRLKSCVFLSRDDI  
AGFLNATQITSENPNELIILLYLNIYNIILDDYTVVFCFLGFKTIFCWLQLR  
CLVNVLEIKDFSGEVAASVWVHNKDLAKFTLLWIPCRFVLMAMEVPVL  
CYNWRTIEGNSGVYDSTPRMIVKVRKTELSILLDQYLLGLDVRNRVRFGR  
PPASQPMOHLFLLPVVNDLSLEMLVPKHPKIPKNVELYLVKSDGVVVAARS  
PLANQSSILKQRMDDIPVLEKNGSNGWDEDEDICNGRDMGDMKNSDVAQ  
VNLQNNISLQDLQDSSGVSKCLSLWLDHDLVGLCFKDDICLKKAVDWCS  
FGRCSCMEADFEKTEFCAWKCKWLEAARMKELEIIEIKTICHTSCALP  
QDSYMEFADFEKTVRIHTQSTFELKMYIEKIGYELOTSDVRIAKNEAIRKVEG  
CQDSFEDPLMAAIIHSSNGLLVWDYDLFPNPKFASFRGVFWFSQSTEGFHRCLP  
IIVTNLNCYQWKLMIASVADANFFLLAFATSTLSTDSWRFSLGIERVYQR  
KGCLISSPDDLLAVINESGOWEPWYNRCLRLHLSQFIDRYVYLDLVKRA  
GTSQSEEDFSWKDIKKNSEARKWLDQPPQWALANDHNGRGYGIEMTEITFRA  
FNVSHDNLVLTGIVLDFELRHSDEFDFCRGSRKCGNVTYPTVEKLAERKDS  
VYVMDPLNDQVTAPOENDEWVOLSCTCGEFCGCKPFLHUALHVALCKLIN  
PLEVVDYCLNLELYTYATFSPVLSAMPEASGPRLPVLPVPPPPPTNVAV  
LITVCSAATLMDPLRENOAKRDLKPTVFFFTFIFSLILSLKLPFLKNPS  
LISPTVTVLMAETLREHMLLYTNACIKYSGDGVYEGSSPKKIRVKNTENT  
LLNGLYIFGLAKQSKISIFGRYPVLSPLDLYLHPVNDSSLETLEVPNSVPS  
IKVELYLEVKSSEGVDPAAACSSPLENGLSKLQRTQBAEYTHSHSVKLERNV  
LKQGVDSNGSIEDEADAALNGSTHYVDREMTNKNSGSDGVEQVITLANNAESR  
LCFDRDLKKTVDWCSIRGQKCVRETEKDYTFECIRKCKWLSQAAREEHLV  
EITKNTPHTCPIGPDNDYVEFAADEIECLIRVQPTLIEELRNWFWFNGDLATS  
EMQAKOEVIKVGWDQDFRVLPLNMAAFHSSNGLVVDWQYKLPNPEFASFRSVF  
WAFQSLTEGHCERPLIVDTKOLGKYPKMLIASGVEADDCYFPLAFATTEVSSD  
TWRFSLGIREKVTQKADICLSRPHDILDVINEGSOEWPWYHMFCLDDICQF  
HYVQDDYLANVLEAGTSEKEEGLFDPVATALLFDEMFFYFSGLCDSSGVRNRGDM  
RYRVMYDAENLFAFESQSLFVPMLEKGLFQVTEPQDEWIVOLSEWSTCGEFLKK  
YPCULHVALEKILNPQVDCYSLDRLYKYAATFSPVEVAWPEASGVPITLFP  
PVILPPNCSGOWKTPRTINDAKVPPSDEELNAIVDLILKDLHLYQVTLGDRKS  
RLVNPVFCWNGIKDGTNGIYEGSNVRYMTVKGTKEFNLHLYQVTLGDRKS  
KGLIIRPDIQAFRYKRLPVVDTSLETLEITTKHPYIKDILYLEVESTVG  
VIDPATCSWSDILGSSGKROKTDHVKLEREDTNRNSESQVHVLVDSTLNKDSISG  
HVSKPCSSWLDLHDLVGLCFKQDGLKAVDCSKAQKCVKRETKADITFEF  
IRWCKWSLAGAARKHGLVEIITYTGHPTCHPVPDEFEFTEFIERAVRYMTQ  
TISELKKWKKIGYELETSDVRLAKEKAKIRVFGWQDQSFEDFPKLMSALCSNGLL  
VDWYDLFPNPFASGCVFWAPQSTEGFHCERPLIVDTKOLGKLYLNIYVNESG  
QWQEPWAYHVNLFHYQSFRVFPSCGLARIRAGTSQKQDEFFSVYMDIEKKEPE  
ARKWLDQPPQNRWALANDHNGRYGIMEINTKALFAVCNAFEAGHVVGVVLLFDEL  
RSKFDKFSRSLNGLNDGYVTEPMDKLEEFFRTYTYIYVTDNNAFQVATADL  
KGECLVOLSDSCCTGDFQRYKFPCLHALAVCKLENPLQYVDDCYTLERLRYAT  
IFSHVPMASWPEASGVRPLLPVIPPSPPTVYVSRVKYVMDCHLNGLYIKYQAD  
GVYEGSTLTKITIMRKTALSRLLDRLYQLFGLDKQSEKIFGKYPVDSPLFTVY  
HFPVNDSSLETLEVPNSVNLFLYEAQPTSDAVVVPVTCSSPLESPSSSKR  
QRITQOEIVDNNMSAGILVVKVNSGLFKCLLPRLWIDDDHMLHGLCLCFKDRDLKK

ADWVCIRRRNRNCIVRETEKEMTYTFECRWKCKWSLRAARAMEEHGLVEITKTYGPHTC  
SHEVNDSEFEFAADEIERVVRIOPTLSIAELKKWKEKTYGELQTSKRDGKLEVIK  
RYGDEQDSFRVMPKILSAHSSNGLLVDMQYDLFPNPOFASPRGVSFOSIEGQF  
HCVGLDVKSLNGKYLKMLTASGDAANKFFPLAFAVTKEVSDTSWRWETKIRE  
KVTQRKDLCLISSPLDIAVAVNPGSLWOEPWAHKKFCLNHLRSOFLGVFRDYNLES  
LVQAGSTNKEEFDSYMDNDIKENPEAKWLDQIPRHWALHSLGRLYGIIEIDRE  
ILFAVCRGFPYCTVAMTGVMFLDELRSFVLSLSIYSSCRKGVYTFEPFMDLEE  
FMDSIPYVITOLERDSFKVSESEKEWIVOLNVSTCTCKRFQSKYKPCCLHALAVFE  
KLKINPQYDECTVYQCKYAAATFSPVDAAMPEDCRVPTLPPPSQOLSPTN"  
complement(join(15674..15817,15904..16008,16075..16196,  
17033..17102))  
/note="similar to ubiquitin-conjugating enzyme E2-17 kd  
spiP3135; similar to ESTs gb|IR84048, gb|T45251,  
gb|AI100555, gb|AA041165, gb|T21697, emb|Z29145,  
gb|T20867"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F22C12.2"  
/protein\_id="AAF24583.1"  
/db\_xref="GI:6692118"  
/translation="MASKRILKELKLDQKDPPTSCSAVAEDMFHWQATIMGPSDSPYS  
GGVFLVTHPEPPDYPKPKVAFETKVEHPNVNSNGSICLDILKEQWSPALTITSKVL  
SICSLIDPNDDPLVFEIAHMYKTDRAKYESTARSWTOKYAMG"  
complement(join(18031..18617,18685..19861))  
/note="putative receptor kinase emb|CAA3040.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F22C12.3"  
/protein\_id="AAF24582.1"  
/db\_xref="GI:6692117"  
/translation="MQIFLFFSLILCFVLISSTOTLEDDKALLHFLSSNSRLHN  
QSSDVCHSWTGVTCNENGDRIVSVRLPAVGNGLIPPTISRSLSLFLSKRNHFTG  
DFPDDTNLKLTHLYLQHNLSGPLLAIFSELKNLKVLDLNNNGSIFPTSLGLT  
SLOVLNANNSFSGEIPNLHLPKLSQNLNNKLIGTIPKSLQFQSSAFSGNNLTER  
KKQRTPLGSLQALFLILSAACVLCVGLSFIMTCFGKTRISGKLRKRDSSPPGN  
KTRSDNVVEEGKIFFGGRNHLFDLDDLSSSEVLGKAGFTTYKVTMEDNSTVV  
KRLVGVVREFQEKIIRHINVEALKEIYKSKDKLAVSYNNGSLPEILH  
GNRGYRVPDLDWDLRIATGAARLAKTHEGRFTHGNIKSSNIFLDSQCYGIGD  
GLTTIMRSLPOTTCITSGYHAPETDTRSTQFSDVYFVGLLETLTGKSPVSQGL  
VPTGENMDLASWIRSVVAKTEVDFMELTSQSGGFEEMVEMLOIQLACVAKQO  
BRPHTAOVLKLIEDIRSVDAE"  
join(20892..20993,21210..21310,21520..21694,21780..21976,  
22063..22162,22282..22387,23017..23183)  
/note="similar to vacuolar ATPase gi|3600058"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F22C12.4"  
/db\_xref="GI:6692094"  
/protein\_id="AAF24559.1"  
/translation="MNDADASTIQOOWVRFTRQAEKANEISISSEEFNLEKLQV  
EAEKKIRQYEKKQVDVRKKIDYSQMLNARIKYLQADDIVNAMKEAALKOLK  
VSOHGFNNHHQYKHLKLDLTVQCLRLKEPALLRCREEDLDIVESMLDDEEYEC  
KKAKVHAPETIVDKDIFLPPAPSDDDPHALSCAGGVVLASRDGKIVCENTLDARLEVA  
FNKLPEFCSGSFLEMCVDPKVALRGQCSLMSDSNFTITKEKLRDAKSMNPTGRRC  
DNGVEKKSMSYSSCKTQGMGSGQGHKGNVMBCEYEG"  
23882..25345  
/note="putative 6-phosphogluconate dehydrogenase  
dbj|BAA22812; similar to ESTs emb|F15308, emb|I5307,  
gb|T21894, gb|AI997770.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="F22C12.5"  
/protein\_id="AAF24560.1"  
/db\_xref="GI:6692095"  
/translation="MESAALSRIGLAGLVAMQNLALNAIEKGFPISVYNTTTSKVD  
TIDRAAEGNLTSGVSPDRFVLSIQPRSLIILVKGAPVDQITDAFSEYMEGDC  
ITDGGNEWQNTERRISEAQGLLYLGMVSGGEGARNGSPGSGFQAYDNIDK  
ILEKVAQVEDGPCVITYEGEGSGNFKVWNGIEYGMOLISEAYDVLKNGVLSNE  
ELAEITFENSGELESFVLEITPDSIFRVKDPGDELVDKILDKMGKCTKWTYQQA  
AELSVAATCSYQAQGNMLLRAKSLKSNLNFGEARLWKGCCIIRAVFLDKKRIALP  
RNPDLASLVDPDEFAQEMVQQAARVVRVGLAVSAGISTPGMCASLAYFDTYRRARLP  
ANLVAQARDLFGAHTYERTDRPGAVHTWTKLARNH"

CDS

CDS

CDS

CDS



```
repeat_region /clone_lib="RPCI-11 human BAC library"
repeat_region /map="17"
repeat_region /chromosome="17"
repeat_region 2. .110
repeat_region /rpt_family="L2"
repeat_region 399. 427
repeat_region /rpt_family="AT-rich"
repeat_region complement(593. .1090)
repeat_region /rpt_family="LMB7"
repeat_region 1477. .1635
repeat_region /rpt_family="MIR"
repeat_region 2171. 2861
repeat_region /rpt_family="LIPAL0"
repeat_region complement(2899. .3103)
repeat_region /rpt_family="MIR"
repeat_region 3384. .3442
repeat_region /rpt_family="G-rich"
repeat_region 4482. .4564
repeat_region /rpt_family="MIR"
repeat_region complement(4565. .4746)
repeat_region /rpt_family="MER69A"
repeat_region 4747. .4784
repeat_region /rpt_family="MIR"
repeat_region 4788. .4966
repeat_region /rpt_family="(TTTTC)n"
repeat_region complement(4967. .5259)
repeat_region /rpt_family="AluSx"
repeat_region 5260. .5357
repeat_region /rpt_family="MIR"
repeat_region 5362. .5418
repeat_region /rpt_family="(CA)n"
repeat_region complement(5523. .5832)
repeat_region /rpt_family="AluSx"
repeat_region 7567. .7732
repeat_region /rpt_family="MER63A"
repeat_region 7751. .7882
repeat_region /rpt_family="MIR"
repeat_region 8008. .8286
repeat_region /rpt_family="L2"
repeat_region 9311. .9531
repeat_region /rpt_family="MIR"
repeat_region complement(9756. .9890)
repeat_region /rpt_family="L2"
repeat_region complement(10044. .10344)
repeat_region /rpt_family="AluSp"
repeat_region complement(10845. .11308)
repeat_region /rpt_family="L2"
repeat_region complement(11335. .11458)
repeat_region /rpt_family="MER69A"
repeat_region complement(11619. .11674)
repeat_region /rpt_family="MER69A"
repeat_region complement(11680. .12175)
repeat_region /rpt_family="L2"
repeat_region 12176. .12204
repeat_region /rpt_family="(TTTTTA)n"
repeat_region complement(12206. .12454)
repeat_region /rpt_family="AluJb"
repeat_region complement(12455. .12843)
repeat_region /rpt_family="L2"
repeat_region 13247. .13529
repeat_region /rpt_family="AluSx"
repeat_region complement(14301. .14370)
repeat_region /rpt_family="MIR"
repeat_region 15962. .16265
repeat_region /rpt_family="AluSg"
repeat_region complement(16706. .16785)
repeat_region /rpt_family="MIR"
repeat_region complement(16804. .16875)
repeat_region /rpt_family="L2"
repeat_region 18053. 18195
repeat_region /rpt_family="HALL"
repeat_region 18328. .19612
repeat_region /rpt_family="HALL"
```

```
repeat_region complement(19638. .19830)
repeat_region /rpt_family="MIR"
repeat_region 19857. .19907
repeat_region /rpt_family="A-rich"
repeat_region 19951. .20217
repeat_region /rpt_family="MLTII"
repeat_region complement(20912. .21095)
repeat_region /rpt_family="MIR"
repeat_region 21171. .21213
repeat_region /rpt_family="AT-rich"
repeat_region 21390. .21534
repeat_region /rpt_family="MER103"
repeat_region complement(22377. .22503)
repeat_region /rpt_family="AluJo/FRAM"
repeat_region 22877. .22973
repeat_region /rpt_family="MIR"
repeat_region 23260. .23297
repeat_region /rpt_family="GA-rich"
repeat_region 24711. .24949
repeat_region /rpt_family="MIR"
repeat_region 25364. .25407
repeat_region /rpt_family="MIR"
repeat_region complement(25408. .25718)
repeat_region /rpt_family="AluSx"
repeat_region complement(25731. .26012)
repeat_region /rpt_family="AluSx"
repeat_region 26919. .26946
repeat_region /rpt_family="AT-rich"
repeat_region 27919. .28060
repeat_region /rpt_family="MIR"
repeat_region complement(29274. .29446)
repeat_region /rpt_family="FRAM"
repeat_region complement(29447. .29548)
repeat_region /rpt_family="MIR"
repeat_region complement(30115. .30239)
repeat_region /rpt_family="HALL"
repeat_region complement(30359. .30444)
repeat_region /rpt_family="LIME"
repeat_region 32686. .32727
repeat_region /rpt_family="(TG)n"
repeat_region 33941. .34089
repeat_region /rpt_family="FRAM"
repeat_region 34120. .34271
repeat_region /rpt_family="L2"
repeat_region 34650. .34683
repeat_region /rpt_family="(CA)n"
repeat_region 34868. .34897
repeat_region /rpt_family="(TATATG)n"
repeat_region complement(35799. .36115)
repeat_region /rpt_family="AluSg"
repeat_region 37092. .37423
repeat_region /rpt_family="L2"
repeat_region 38231. .38278
repeat_region /rpt_family="CT-rich"
repeat_region complement(38610. .38723)
repeat_region /rpt_family="AluSg/x"
repeat_region complement(38749. .38823)
repeat_region /rpt_family="MER63C"
repeat_region complement(38824. .39178)
repeat_region /rpt_family="LIPB1"
repeat_region 39180. .39214
repeat_region /rpt_family="(TA)n"
repeat_region complement(39215. .39605)
repeat_region /rpt_family="MER63C"
repeat_region 39618. .39667
repeat_region /rpt_family="(TA)n"
repeat_region 40234. .40269
```

Query Match 16.7%; Score 38.2; DB 10; Length 170453;  
Best Local Similarity 53.7%; Pred. No. 0.64;  
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 68 agaataggccttgctgattgctgttaatggcacaactgctgcaacttattgttga 127

```

Db 1879 AAAAGAGGCTCTCTCTATTATCTGTGAAGAGCAAAATCTCTCACCTTGCCCTGG 1820
QY 128 aaagggtcccaattccggttaacaaaggaacattccagtggtatgggccaataaga 187
Db 1819 GAAGGAAAAGAGAGATATTGTGTGACAGAGTCCCTTATGAGCAAGTAATGTCATGAGA 1760
QY 188 cgaagcaaacagggaaggaacattca 214
Db 1759 GTATTTCACAGCAAGGCCCTCTCA 1733

RESULT 13
AC015724/c
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone RP11-45J24 map 17, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
ACCESSION
AC015724
VERSION
AC015724.4 GI:7329270
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human
ORGANISM - Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184799)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Gaigani,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGuck,K., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,K.R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6751809.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1414
Center clone name: 45_J24
----- Summary Statistics
Sequencing vector: M13; M7815; 96% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
3.61918146119Chemistry: Dye-terminator Big Dye; 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 164567 bases at least Q40
Consensus quality: 174814 bases at least Q30
Consensus quality: 179219 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 183199; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality cov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces

```

\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 2143: contig of 2143 bp in length
* 2144 2243: gap of 100 bp
* 2244 6178: contig of 3935 bp in length
* 6179 6278: gap of 100 bp
* 6279 10043: contig of 3765 bp in length
* 10044 10143: gap of 100 bp
* 10144 12960: contig of 2817 bp in length
* 12961 13060: gap of 100 bp
* 13061 17800: contig of 4740 bp in length
* 17801 17900: gap of 100 bp
* 17901 22268: contig of 4368 bp in length
* 22269 22368: gap of 100 bp
* 22369 27281: contig of 4913 bp in length
* 27282 27381: gap of 100 bp
* 27382 32880: contig of 5499 bp in length
* 32881 32980: gap of 100 bp
* 32981 41040: contig of 8060 bp in length
* 41041 41140: gap of 100 bp
* 41141 48535: contig of 7395 bp in length
* 48536 48635: gap of 100 bp
* 48636 56130: contig of 7495 bp in length
* 56131 56230: gap of 100 bp
* 56231 71938: contig of 15708 bp in length
* 71939 72038: gap of 100 bp
* 72039 86256: contig of 14218 bp in length
* 86257 86356: gap of 100 bp
* 86357 100509: contig of 14153 bp in length
* 100510 100609: gap of 100 bp
* 100610 122739: contig of 22130 bp in length
* 122740 122839: gap of 100 bp
* 122840 144885: contig of 22046 bp in length
* 144886 144985: gap of 100 bp
* 144986 184799: contig of 39814 bp in length.

```

## FEATURES

## source

```

1. 184799
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone_lib="RP11-45J24"
/clone_lib="RP11-11 Human Male BAC"
1. 2143
/note="assembly_fragment"
2244. 6178
/note="assembly_fragment"
6279. 10043
/note="assembly_fragment"
10144. 12960
/note="assembly_fragment"
13061. 17800
/note="assembly_fragment"
17901. 22268
/note="assembly_fragment"
22369. 27281
/note="assembly_fragment"
27382. 32880
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
32981. 41040
/note="assembly_fragment"
clone_end:17
vector_side:right"
41141. 48535
/note="assembly_fragment"
48636. 56130
/note="assembly_fragment"

```

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

misc_feature	56231.	.71938	/note="assembly_fragment"	
misc_feature	72039.	.86256	/note="assembly_fragment"	
misc_feature	86357.	.100509	/note="assembly_fragment"	
misc_feature	100610.	.122739	/note="assembly_fragment"	
misc_feature	122840.	.144885	/note="assembly_fragment"	
misc_feature	144986.	.184799	/note="assembly_fragment"	
BASE COUNT	52620	a	39686	c 39593 g 51292 t 1608 others
ORIGIN				

Query Match	16.7%	Score 38.2;	DB 38;	Length 184799;
Best Local Similarity	53.7%;	Pred. No. 0.64;		
Matches 79;	Conservative 0;	Mismatches 68;	Indels 0;	Gaps 0;
QY 68	agaataggcctgcgtgattggctgtgtaattggcgaaatcggcgacctcaatatgtctga	127		
Db 37009	AAAAGGGCCTCTCATTATCTGTGTAAGAGCAAAAATCTCTCTACCCCTTGCTGG	36950		
QY 128	aaagggtctcccaattccggtttaacaacggaacattcccaaggttatggggccataaga	187		
Db 36949	GAAGGAAAAGAGAGTATTGTGGACAAGAGTCCCTATGAGCAAAAGTAATTGGCAATGACA	36890		
QY 188	cgaagcacaaccaggaaagaaaccttca	214		
Db 36889	GTATTTCCAACAGAAAGGCCCTCTCA	36863		

RESULT	14
LDL130772	
LOCUS	LDL130772 3158 bp mRNA PLN 17-NOV-1999
DEFINITION	Laminaria digitata mRNA 6-phosphogluconate dehydrogenase.
ACCESSION	AJ130772
VERSION	AJ130772.1 GI:6453556
KEYWORDS	6-phosphogluconate dehydrogenase; 6PGDH gene.
SOURCE	Laminaria digitata.
ORGANISM	Laminaria digitata.
REFERENCE	Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group; Phaeophyceae; Laminariales; Laminariaceae; Laminaria. 1 (bases 1 to 3158)
AUTHORS	Crepineau,F., Roscoe,T., Kaas,R., Kloareg,B. and Boyen,C.
TITLE	Alteration of generations in Laminaria digitata: a comparison of the gametophyte and sporophyte physiology by an EST strategy unpublished
JOURNAL	2 (bases 1 to 3158)
REFERENCE	Moulin,P., Crepineau,F., Kloareg,B. and Boyen,C.
AUTHORS	

\*\*\*\*\*  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

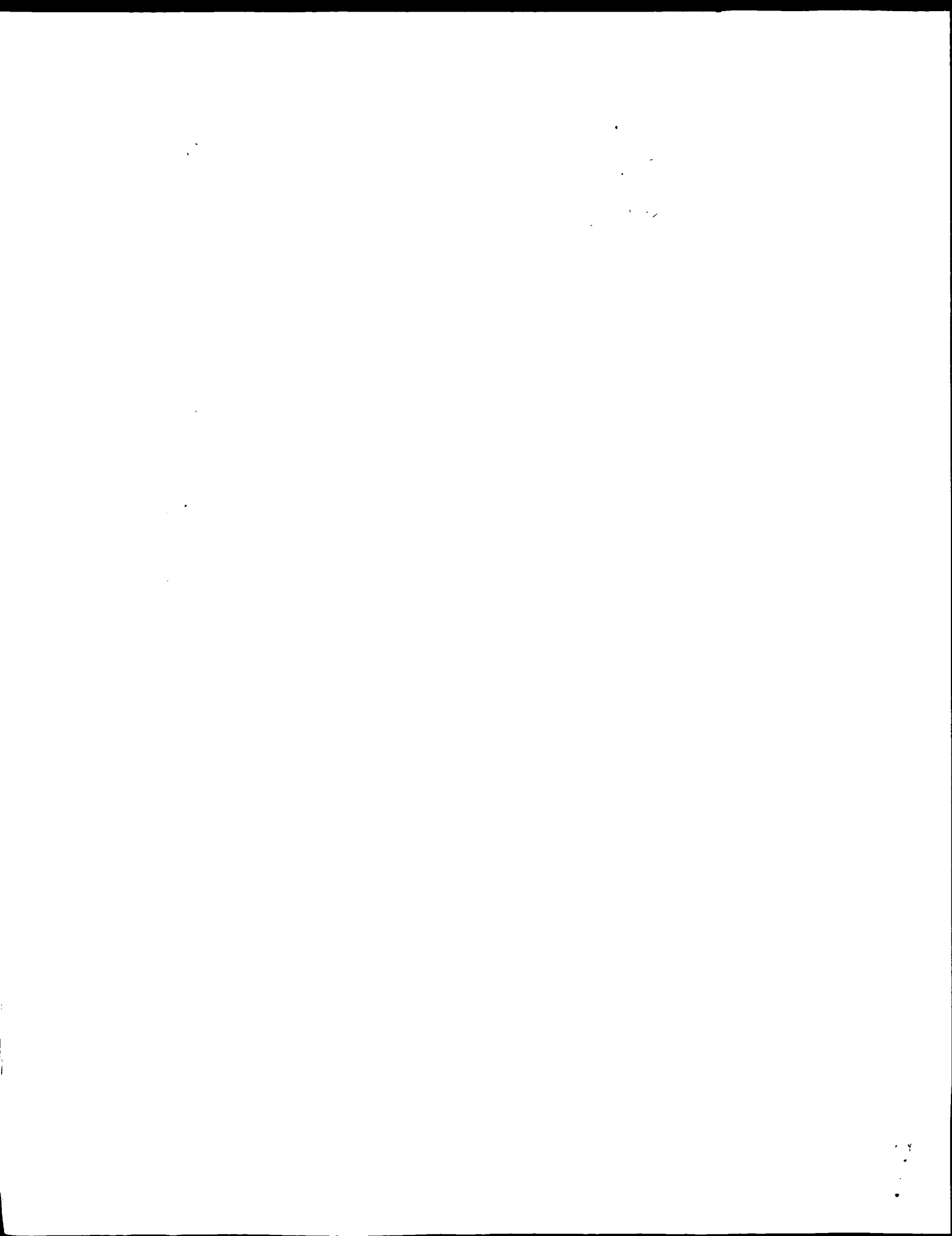
```

1..3138
/organism="Laminaria digitata"
/db_xref="taxon:80365"
/dev_stage="sporophyte"
68..1660
/gene="6PGDH"
68..1660
/gene="6PGDH"
/EC_number="1.1.1.44"
/codon_start=1
/product="6-phosphogluconate dehydrogenase"

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:01 ; Search time 4352.3 Seconds  
(without alignments)  
582.439 Million cell updates/sec

Title: US-09-300-482-14  
Perfect score: 410  
Sequence: 1 cccacgcgtccgggtcatg.....gcaacggccgccttgatg 410

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: gb\_estl.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estpl1.\*  
71: em\_estpl2.\*  
72: em\_estpl3.\*  
73: em\_estpl4.\*  
74: em\_estpl5.\*  
75: em\_estro1.\*  
76: em\_estro2.\*  
77: em\_estro3.\*  
78: em\_estro4.\*  
79: em\_estro5.\*  
80: em\_estro6.\*  
81: em\_estro7.\*  
82: em\_estro8.\*  
83: em\_estro9.\*  
84: em\_estro10.\*  
85: em\_estro11.\*  
86: em\_estro12.\*  
87: em\_estro13.\*  
88: gb\_gss1.\*  
89: gb\_gss2.\*  
90: gb\_gss3.\*  
91: gb\_gss4.\*  
92: em\_gss1.\*  
93: em\_gss2.\*  
94: em\_gss3.\*  
95: em\_gss4.\*  
96: gb\_gss5.\*  
97: gb\_gss6.\*  
98: gb\_gss7.\*  
99: gb\_gss8.\*  
100: gb\_gss9.\*  
101: em\_gss5.\*  
102: em\_gss6.\*  
103: em\_gss7.\*  
104: em\_gss8.\*  
105: em\_gss9.\*  
106: em\_gss10.\*  
107: em\_gss11.\*  
108: gb\_gss10.\*  
109: gb\_gss11.\*  
110: em\_gss12.\*  
111: gb\_gss12.\*  
112: gb\_gss13.\*  
113: gb\_gss14.\*  
114: gb\_gss15.\*  
115: gb\_gss16.\*  
116: gb\_gss17.\*

117: gb\_gss18:\*  
 118: gb\_gss19:\*  
 119: em\_gss13:\*  
 120: gb\_gss20:\*  
 121: gb\_gss21:\*  
 122: gb\_gss22:\*  
 123: gb\_gss23:\*  
 124: gb\_gss24:\*  
 125: em\_gss14:\*  
 126: em\_gss15:\*  
 127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	346.8	84.6	530	35	BE363077 DGL_9_F06
2	323.8	79.0	729	34	BE216705 HV_CED001
3	323.8	79.0	886	35	BE413061 MCG015.E0
4	320.6	78.2	466	35	BE471136 WHE0284.C
5	319	77.8	547	36	BE493869 WHE1276.F
6	314.6	76.7	519	23	AW563648 LGL_245.A
7	313.6	76.5	516	35	BE363025 DGL_9_A10
8	295.2	72.0	743	35	BE414610 SCU011.F0
9	290.4	70.8	456	36	D24970 RICR2869A.R
10	290.2	70.8	472	36	BE497862 WHE0957.D
11	288.6	70.4	474	36	BE497880 WHE0957.B
12	288.6	70.4	512	35	BE425231 WHE0312.G
13	281.4	68.6	454	36	BE495944 WHE1259.B
14	277.4	67.7	781	35	BE414643 SCU011.H1
15	269	65.6	463	21	AW286879 LGL_222.F
16	268	65.4	462	21	AW286839 LGL_222.B
17	267	65.1	461	21	AW286858 LGL_222.D
18	266	64.9	460	23	AW677090 DGL_4.B11
19	264.4	64.5	362	35	BE403238 WHE0430.A
20	263.6	64.3	528	19	AV550009 AV550009
21	262	63.9	530	19	AV526953 AV526953
22	243.8	59.5	457	36	BE494267 WHE1252.B
23	241.4	58.9	374	15	AU068515 AU068515
24	241.2	58.8	638	21	AW254720 ML1006.pe
25	237	57.8	532	24	AW719787 LjNEST10A
26	231.6	56.5	561	32	BE022117 sm67f12.Y
27	230	56.1	504	35	BE450814 EST401701
28	229	55.9	573	23	AW586409 EST318032
29	228.4	55.7	479	14	AL365745 MBLA02B10
30	228.4	55.7	481	14	AL365743 MBLA02B09
31	228.4	55.7	653	24	AW695821 NF099E09S
32	227.4	55.5	630	23	AW688619 NF009F08S
33	226.8	55.3	499	32	BE020243 sm42404.Y
34	226.8	55.3	606	21	AW299197 EST306007
35	225.2	54.9	578	36	BE496464 NXC1_018
36	224.6	54.8	564	23	AW585410 EST317033
37	222	54.1	498	35	BE341175 EST345276
38	221.2	54.0	401	35	BE428880 WTD011.H0
39	214.8	52.4	548	14	AL368600 MBLA25FE04
40	210.4	51.3	394	18	AV425788 AV425788
41	210.4	51.3	430	14	AL368071 MBLA22A07
42	209.4	51.1	420	14	AL382847 MBLA10C10
43	206	50.2	592	23	AW560451 EST315499
44	202.4	49.3	430	18	AV406682 AV406682
45	200.4	48.9	358	36	C74420 C74420 Rice

## ALIGNMENTS

RESULT 1  
 BE363077

LOCUS BE363077 530 bp mRNA EST 20-JUL-2000  
 DEFINITION DGL\_9\_F06.bl\_A002 Dark Grown 1 (DGL1) Sorghum bicolor cDNA, mRNA sequence.  
 ACCESSION BE363077  
 VERSION BE363077.1 GI:9304634  
 KEYWORDS EST.  
 SOURCE sorghum.  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poales; Poaceae; Sorghum.  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.  
 1 (bases 1 to 530)  
 REFERENCE Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.  
 An EST database from Sorghum: dark-grown seedlings  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
 Seq primer: JEN REV  
 High quality sequence stop: 453  
 POLYA-No. Location/Qualifiers  
 1..530  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGL1)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site.1: XhoI; Site.2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 126 a 132 c 164 g 108 t  
 ORIGIN  
 Query Match 84.6%; Score 346.8; DB 35; Length 530;  
 Best Local Similarity 91.7%; Pred. No. 9.4e-81;  
 Matches 366; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
 QY 12 gcggtcatgggcagaaaccttgcctcaacattgcagagaaagggtcccatctctgtg 71  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 91 GCGGTCTATGGGCGAGAACCTTGCCCTCAACATTGCAGAGAAAGGCTCCCATCTCTGTC 150  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 72 tacaacagagacaacctccaaaggtgagcagagacgtgcagcgtgccaaaggcagaagaaac 131  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 151 TACAACAGGACACACCTCCCAAGGTTGATGAGACCGTGCACGCGTCCCAAGGACGAAAGAAAC 210  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 132 cttcccgctacggtctccatgaccctgcctcttcttgtaagtcattcagagcagcag 191  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 211 CTTCCCGTGTACGGCTCCATGATGACCCCTGCATCTCTTGTGAATCCATTCAGAACGCGGT 270  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 192 gtggtgatcatgctgtcaagcgccgagcgccgagtcagtcagcagtcagtcagtcagtcag 251  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 271 GTTGTATCATGTTCTGTAAGGCTGTGTGACCAAGTTGACCAAGCATTCGACGCTCGCA 330  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 252 gctcaacttgagcagggcgactgcatcatcgatgggggagaaagagtggtacagaaacacg 311  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 331 GCTCACTTGGAGAGAGGCTGACTGTATCATGTATGAGGAGGAATGAGTGGTATGAGAACT 390  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 312 gagagagggagagggccatggagagcgcgccctnctgtatcttggcattgggtgtctct 371  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 391 GAGAGAGGAGAGAGGCCATGGAGAGCGTGGCTCTCTCTATCTTGAATGGCGTCTCT 450  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 372 ggaggaagaggggtgcccgcgaacggcccgctcttgatg 410  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 451 GGAGGAGAGGAGGCTGCCCAACGAGACCATCTCTTGATG 489

358 AGAAGGAGGAGGAGGATGGAGGAGCGTGGAGTCTCTACCTCGTATGTTGTTTC 417

371 tggaggaaaggaggtgcccgcgaacgcccgtctctgtg 410

418 TGGAGGAGGAGGAGGTGCCGCCCATGGCCGTCATCATG 457

RESULT 3

BE216705 729 bp mRNA EST 25-JUL-2000

LOCUS HV\_CEB0011E16f Hordeum vulgare seedling green leaf EST library

DEFINITION HVCNDA0005 (Erysiphe infected & control) Hordeum vulgare cDNA clone

ACCESSION BE216705

VERSION BE216705.1 GI:8904317

KEYWORDS EST.

SOURCE Hordeum vulgare

ORGANISM Hordeum vulgare

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Bequm, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo, T., Sasaki, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.

TITLE Development of a genetically and physically anchored EST resource for barley genomics

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: AATTACCTCTCAATAAGG

High quality sequence start: 20

High quality sequence stop: 678.

FEATURES

Location/Qualifiers

1..729

/organism="Hordeum vulgare"

/cultivar="C116151 (Mla6)"

/db\_xref="taxon:4513"

/clone\_lib="HV\_CEB0011E16f"

/clone\_lib="Hordeum vulgare seedling green leaf EST library HVCNDA0005 (Erysiphe infected & control)"

/tissue\_type="seedling green leaf"

/lab\_host="SOLR"

/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 166 a 206 c 211 g 146 t

ORIGIN

Query Match 79.0%; Score 323.8; DB 34; Length 729;

Best Local Similarity 88.0%; Pred. No. 1e-74;

Matches 352; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 11 cgcggtcatggtggcagaaaccttgcctcaacattgcagagaaagggttccccatctctgt 70

Db 58 CGCGTCTATGGGAGAGACCTTGCCTCAACATTGCAGAGAAAGGCTTCCCATCTCTGT 117

Qy 71 gtacaacaggagcaacctccaagggtggagagacccgtgcagcgtgccaaagcagaagaaa 130

Db 118 CTACAACAGGACGACATCCAAAGGTCGAGAGACCGTCCAGCGCCCAAGTGTAGAAGGAAA 177

Qy 131 ccttcctctacggtctccatgaccccgctctcttgaagtccattcagaagccacg 190

Db 178 CTTCTCTCTACGCTTTCATGACCTGATCTCTTTGTCACTCCATTCAGAACGCCACG 237

Qy 191 ggtggtatcatgctgcgtcaaggccgcccgttccttgaagtccattcagaagccacg 250

Db 238 TGTCTCTATCATGCTTGTCAAGGCGGCTGCTCCGCTTTCAGCAGACCATCGCCGCTCGC 297

Qy 251 agctcaattgagcagggcagctcatcatgatggggggaacagtggttacgagaacac 310

Db 298 AGCACACCTGGAGAGGCGACTGTCATGTTGACGGAGGAACAGTGTATGAGAACAC 357

Qy 311 ggagagggagagagccatggagagcgccgctcctctatcttggcatgggtgtctc 370

358 AGAAGGAGGAGGAGGATGGAGGAGCGTGGAGTCTCTACCTCGTATGTTGTTTC 417

371 tggaggaaaggaggtgcccgcgaacgcccgtctctgtg 410

418 TGGAGGAGGAGGAGGTGCCGCCCATGGCCGTCATCATG 457

RESULT 3

BE413061 886 bp mRNA EST 24-JUL-2000

LOCUS MCG015.E0590625 ITEC MCG Barley Leaf/Culm Library Hordeum vulgare

DEFINITION cDNA clone MCG015.E05, mRNA sequence.

ACCESSION BE413061

VERSION BE413061.1 GI:9410907

KEYWORDS EST.

SOURCE Hordeum vulgare

ORGANISM Hordeum vulgare

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.

TITLE International Triticeae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae

JOURNAL Unpublished (2000)

COMMENT Contact: Graner A

Institute for Plant Genetics & Crop Plant Research

Corrensstr. 3, D-06466 Gatersleben GERMANY

Tel: 49 39482 5521

Fax: 49 39482 5137

Email: a.graner@ipk-gatersleben.de

International Triticeae EST Cooperative (ITEC)

http://wheat.pw.usda.gov/genome.

FEATURES

Location/Qualifiers

1..886

/organism="Hordeum vulgare"

/db\_xref="taxon:4513"

/clone\_lib="MCG015.E05"

/clone\_lib="ITEC MCG Barley Leaf/Culm Library"

/tissue\_type="leaf/culm"

/dev\_stage="etiolated"

BASE COUNT 202 a 224 c 258 g 186 t 16 others

ORIGIN

Query Match 79.0%; Score 323.8; DB 35; Length 886;

Best Local Similarity 88.0%; Pred. No. 1e-74;

Matches 352; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 11 cgcggtcatggtggcagaaaccttgcctcaacattgcagagaaagggttccccatctctgt 70

Db 47 CGCGTCTATGGGAGAGACCTTGCCTCAACATTGCAGAGAAAGGCTTCCCATCTCTGT 106

Qy 71 gtacaacaggagcaacctccaagggtggagagacccgtgcagcgtgccaaagcagaagaaa 130

Db 107 CTACAACAGGACGACATCCAAAGGTCGAGAGACCGTCCAGCGCCCAAGTGTAGAAGGAAA 166

Qy 131 ccttcctctacggtctccatgaccccgctctcttgaagtccattcagaagccacg 190

Db 167 CTTCTCTCTACGCTTTCATGACCTGATCTCTTTGTCACTCCATTCAGAACGCCACG 226

Qy 191 ggtggtatcatgctgcgtcaaggccgcccgttccttgaagtccattcagaagccacg 250

Db 227 TGTCTCTATCATGCTTGTCAAGGCGGCTGCTCCGCTTTCAGCAGACCATCGCCGCTCGC 286

Qy 251 agctcaattgagcagggcagctcatcatcgatggggggaacagtggttacgagaacac 310

Db 287 AGCACACCTGGAGAGGCGACTGTCATGTTGACGGAGGAACAGTGTATGAGAACAC 346



pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

```

BASE COUNT      121 a   149 c   160 g   117 t
ORIGIN

Query Match      77.8%; Score 319; DB 36; Length 547;
Best Local Similarity 87.2%; Pred. No. 1.8e-73;
Matches 349; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 11 cgcggtcagggcagaaaccttgcctcaacattgcagaaagggttcccatctctgt 70
Db 48 cgcggtcagggcagaaaccttgcctcaacattgcagaaagggttcccatctctgt 107

QY 71 gtacaacaggacaaaccttccaaaggttgagacagcgtgcagcgtgccaaaggcagaagaaa 130
Db 108 ctacaacaggacaaaccttccaaaggttgagacagcgtgcagcgtgccaaaggcagaagaaa 167

QY 131 ccttcctctacggttcctcattgaccccgctcctttgtgaagtcattcagaagccagc 190
Db 168 ccttcctctacggttcctcattgaccccgctcctttgtgaagtcattcagaagccagc 227

QY 191 ggtgtgatcatgctcgtcaaggccgcgcgcagttgacacagaccatcgacgctcgc 250
Db 228 cgtcgtcatcatgctcgtcaaggccgcgcgcagttgacacagaccatcgacgctcgc 287

QY 251 agtcaacttgagcagggcgactgcattgcattgagggggaacagtggttacgagaacac 310
Db 288 tgcacaccctggagcaaggtgactgcattgcattgagggggaacagtggttacgagaacac 347

QY 311 ggagagggaggaagccatggaggagcgcgccctnctgtatttggcattgggtgtctc 370
Db 348 ggaaggagggaggaagccatggaggagcgcgccctnctgtatttggcattgggtgtctc 407

QY 371 tggagggaaggaggtgcccccaacgcgcgcctccttgatg 410
Db 408 cggaggagggagggcgccccccatggcccgctccatgatg 447

RESULT 6
AW563648      519 bp      mRNA      EST      19-JUL-2000
LOCUS      LG1_245_A11.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
DEFINITION      sequence.
ACCESSION      AW563648
VERSION      AW563648.1 GI:7217526
KEYWORDS      EST.
SOURCE      sorghum.
ORGANISM      Sorghum bicolor
REFERENCE      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS      Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
TITLE      1 (bases 1 to 519)
JOURNAL      Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
COMMENT      An EST database from Sorghum: light-grown seedlings
              Unpublished (2000)
              Contact: Cordonnier-Pratt MM
              Department of Botany
              The University of Georgia
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 542 1805
              Email: mmpratt@uga.edu
              Sequences have been trimmed to exclude PolyA, vector and regions
              below Phred quality 16. The threshold for highest quality sequence
              is 20.
              Seq primer: JEN REV
              POLYA-No.
              Location/Qualifiers
              i. 519
              /organism="Sorghum bicolor"

BASE COUNT      118 a   137 c   157 g   107 t
ORIGIN

Query Match      76.7%; Score 314.6; DB 23; Length 519;
Best Local Similarity 91.7%; Pred. No. 2.5e-72;
Matches 332; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 12 gcggtcatggggcagaaaccttgcctcaacattgcagaaagggttcccatctctgtg 71
Db 158 GCGGTCAATGGGGCAGAACCTTGCCTCAACATTGCAGAAAGGCTTCCCATCTCTGTC 217

QY 72 tacaacaggacaaaccttccaaaggttgagacagcgtgcagcgtgccaaaggcagaagaaaac 131
Db 218 TACAACAGGACAAACCTCCAAGGTTGATGAGACCGTGCAGCGTCCCAAGGCAGAGGAAC 277

QY 132 cttccgctctacggttcctcattgaccccgctcctttgtgaagtcattcagaagccagc 191
Db 278 CTTCCTGTGTACGGCTTCCATGACCTGCTATCCTTTGTGAACCTCCATTCAGAAGCCGCT 337

QY 192 gttgtgatcatgctcgtcaaggccgcgcgcagttgacacagaccatcgacgctcgc 251
Db 338 GTTGTCAATCATGCTTGTCAAGGCTGTTGACCACTGTTGACCACTGTTGACCACTGTTGACCACT 397

QY 252 gctcaacttgagcagggcgactgcattgcattgcattgcattgcattgcattgcattgcatt 311
Db 398 GCTCACTTGGAGCAGGGTGACTGCTATCATTTGATGAGGGAATGAGTGTATGAGAACA 457

QY 312 gagagagaggaagccatggaggagcgcgccctnctgtatttggcattgggtgtctct 371
Db 458 GAGAGGAGGAGGAAGGCCATGGAGGAGCGTGGCCCTCTCTATCTTTGGAATGGGCGCTCTCT 517

QY 372 gg 373
Db 518 GG 519

RESULT 7
BE363025      516 bp      mRNA      EST      20-JUL-2000
LOCUS      DGL_9_A10.b1_A002 Dark Grown 1 (DGL) Sorghum bicolor cDNA, mRNA
DEFINITION      sequence.
ACCESSION      BE363025
VERSION      BE363025.1 GI:9304582
KEYWORDS      EST.
SOURCE      sorghum.
ORGANISM      Sorghum bicolor
REFERENCE      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS      Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
TITLE      1 (bases 1 to 516)
JOURNAL      Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
COMMENT      An EST database from Sorghum: dark-grown seedlings
              Unpublished (2000)
              Contact: Cordonnier-Pratt MM
              Department of Botany
              The University of Georgia
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 542 1805
              Email: mmpratt@uga.edu
              Sequences have been trimmed to exclude PolyA, vector and regions
              below Phred quality 16. The threshold for highest quality sequence
              is 20.
              Seq primer: JEN REV
              High quality sequence stop: 516
  
```

POLYA=NO. Location/Qualifiers

FEATURES source

1. 516  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4538"  
/clone\_lib="Dark Grown 1 (DGL)"  
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 117 a 136 c 155 g 108 t

ORIGIN

Query Match 76.5%; Score 313.6; DB 35; Length 516;  
Best Local Similarity 91.7%; Pred. No. 4.5e-72;  
Matches 331; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 12 gcggtcatggggcagaacattgcccctcaacattgcagagaaagggttccccattctgtg 71  
|||||  
Db 156 GCGGTCTATGGGCGAGAACCTTGCCTCAACATTGCAGAGAAAGGCTTCCCATCTCTGTC 215  
|||||

Qy 72 tacacaggacaaacctccagggtgagcagaccgtgcagcgtgccaaaggcagaagaac 131  
|||||  
Db 216 TACAACAGGACAACTCCAAAGTTGATGAGACCGTGCAGCGTGCCAAAGGCGAGAAAC 275  
|||||

Qy 132 ctccccctacggcttccatgaccgcgcctcttggagtcattcagaagccacgg 191  
|||||  
Db 276 CTTCTGTGTACGGCTTCCATGACCCCTGCACTCTTGTGAACATCCATTCAGAGCCGCGT 335  
|||||

Qy 192 gtgtgatcatgctcgaaggccgcgcgcctgtgacagaccattcgagcgtcgca 251  
|||||  
Db 336 GTTGTATCATGCTTGTCAAGCTGGTGACACAGTTGACACACCATTTGGACGCTCGCA 395  
|||||

Qy 252 gctcaattgagcaggcgactgcatactcagatgagggggaacagatgagcagaacag 311  
|||||  
Db 396 GCTCACTGGAGCAGGCTGATGTATTCATTGATGGAGGAATGAGTGTATGAGAACACT 455  
|||||

Qy 312 gagaggaggaagccatgagagcgcgcctctctatcttgcatggtgtctct 371  
|||||  
Db 456 GAGAGGAGGAGAGGCAATGGAGGAGCGTGCGCTCTCTATCTTGAATGGCGTCTCT 515  
|||||

Qy 372 g 372  
|  
Db 516 G 516

RESULT 8  
BE414610 743 bp mRNA EST 24-JUL-2000  
LOCUS SCU011.F01.R90714 ITPC SCU Wheat Endosperm Library Triticum  
DEFINITION aestivum CDNA clone SCU011.F01, mRNA sequence.

ACCESSION BE414610  
VERSION BE414610.1 GI:9412456  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.

REFERENCE 1 (bases 1 to 743)  
AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P., Hermann, R.G., Holton, T., Jacquemien, J.M., Jia, J., Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrells, M., Warburton, M., and Wenzel, G.  
TITLE International Triticaceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticaceae  
JOURNAL Unpublished (2000)  
COMMENT Contact: Holton T  
Centre for Plant Conservation Genetics, Southern Cross University  
PO Box 157, Lismore NSW 2480 AUSTRALIA  
Tel: 61 2 6620 3409

Fax: 61 2 6622 2080  
Email: tholton@scu.edu.au  
International Triticaceae EST Cooperative (ITEC)  
http://wheat.pw.usda.gov/genome.  
Location/Qualifiers

FEATURES source

1. 743  
/organism="Triticum aestivum"  
/cultivar="Wyuna"  
/db\_xref="taxon:4565"  
/clone="SCU011.F01"  
/clone\_lib="ITEC SCU Wheat Endosperm Library"  
/tissue\_type="endosperm"  
/note="Vector: Bluescript II SK(-)"

BASE COUNT 174 a 189 c 207 g 155 t 18 others

ORIGIN

Query Match 72.0%; Score 295.2; DB 35; Length 743;  
Best Local Similarity 83.9%; Pred. No. 3.1e-67;  
Matches 344; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

Qy 2 ccacggtccggtcattggggcagaaaccttgccctcaacattgcagagaagggttccc 61  
|||||  
Db 93 CCCCCGGGTGCGAGAAATTCGGCAGAGTCCGCCCTCAACATTGCGGAGAAAGGTTCCC 152  
|||||

Qy 62 catctctgtgtacaacaggaacacctccaaggtgagcagaccgtgcagcgtgccaaagg 121  
|||||  
Db 153 CATCTCTGTCTACACAGGACCCCTCCAAAGTGCATGAGACTGTTACGCGCGCAAGCT 212  
|||||

Qy 122 agaagaaaccttccgtctacgttccatgaccccgctccttcttgagagtcattca 181  
|||||  
Db 213 AGAAGGAAACCTTCTCTACGGTTTCATGACCTGCATCTTCGTCAACTCATTCA 272  
|||||

Qy 182 gaagcagcgggtggtgatcatgctctcgaagcgcgcgcagtcagcagaccatgcg 241  
|||||  
Db 273 GAAGCCACGTGTCATCATGCTTGTCAAGCCGGTGCCCGGTTGACCAACCATTTGC 332  
|||||

Qy 242 gacgtcgcagctcacttggagcagggcagctgcatactcagatggggggaacg-agtgg 300  
|||||  
Db 333 AACGCTCGCAGCACACCTTGGAGCAGGCGACTGTCATCTGTATGAGGAGGAACGAAGTGT 392  
|||||

Qy 301 acgagaacacagagaggaggaggaagccatggagagcgcgcgcctctctgtatttgga 360  
|||||  
Db 393 ATGAGAACACGGAAGAGGAGGAGGATGGAGGCGTGGAGCTCTTACCTCGGGA 452  
|||||

Qy 361 tgggtgtcttgagagaaagggtgccgcaacgcgcgcctctctgatg 410  
|||||  
Db 453 TGGGTGTTTTCCGGAGGAGGAAGANGGTGCCCGCATGGCCCATCCATGATG 502  
|||||

RESULT 9  
D24970 456 bp mRNA EST 08-JUL-1999  
LOCUS R1C2869A Rice root Oryza sativa cDNA clone R2869\_1A, mRNA  
DEFINITION sequence.  
ACCESSION D24970  
VERSION D24970.1 GI:428814  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

REFERENCE 1 (bases 1 to 456)  
AUTHORS Minobe, Y. and Sasaki, T.  
TITLE Rice cDNA from root  
JOURNAL Unpublished (1995)  
COMMENT Contact: Yuzo Minobe  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441



Fax: 0298-38-7468  
Email: minob@rcs0.riken.go.jp  
PROJECT = "RGP"

## FEATURES

source  
Location/Qualifiers  
1. .456  
/organism="Oryza sativa"  
/strain="Nipponbare, sub\_species Japonica"  
/db\_xref="taxon:4530"  
/clone="R2869\_1A"  
/clone\_lib="Rice root"  
/note="Prepared from seedling root."  
107 a 104 c 130 g 112 t 3 others

BASE COUNT 107 a 104 c 130 g 112 t 3 others

Query Match 70.8%; Score 290.4; DB 36; Length 456;  
Best Local Similarity 85.6%; Pred. No. 5.2e-66;  
Matches 332; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

QY 12 gcggctatggggcagacaccttgcctcaacattgcagagaagggttccccatctctgtg 71  
|||||  
Db 70 GCGGTCATGGGCGCAGACCTTGCCTCAACATTGCAGAGAAAGGTTCCCTATCTCTGTC 129  
|||||  
QY 72 tacaacaggaacaacctcaagtggaagcagaccgtgacggtgcaaggcagagaagaaac 131  
|||||  
Db 130 TACAACAGGACGACTTCTAAGGTTGATGAGACCGTTCAGCGCCAGGTAGAGGAAC 189  
|||||  
QY 132 cttcccgctacggtccatccatgaccccgctccttcttgagtcattccagaagccacg 191  
|||||  
Db 190 CTTCTGTGTACGGGTTTCATGACCTGCAATCTTGTGAACCTCAATCAGAGCCACGT 249  
|||||  
QY 192 gtggtgatcatgctgcagagccgagccgagcagtgagtgagtgagtgagtgagtgag 251  
|||||  
Db 250 GTTGTATCATGCTGTCAAGGCTGGTGACCACTGAGGACGAGACCATTTGCAACTTGTGCA 309  
|||||  
QY 252 gctcaactggagcgggagcagtcagtcagtgagtgagtgagtgagtgagtgagtgag 311  
|||||  
Db 310 GCACACTTGGAGCAGGCTGACTGATTATTGATGAGGAAATGATGTTGACGAGACACT 369  
|||||  
QY 312 gag 371  
|||||  
Db 370 GAGAGAGGAG 429  
|||||  
QY 372 ggaggaag 399  
|||||  
Db 430 GGA-GAGAGGAGGTTCCCGAATGGGCC 456

## RESULT 10

BE497862 472 bp mRNA EST 04-AUG-2000  
LOCUS  
DEFINITION WHE0957\_D04\_H072S Wheat pre-anthesis spike cDNA library Triticum aestivum CDNA clone WHE0957\_D04\_H07, mRNA sequence.  
ACCESSION BE497862  
VERSION BE497862.1 GI:9696479  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.  
REFERENCE 1 (bases 1 to 472)  
AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.  
TITLE The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818

Email: oandersn@pw.usda.gov  
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.

## FEATURES

source  
Location/Qualifiers  
1. .472  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE0957\_D04\_H07"  
/clone\_lib="Wheat pre-anthesis spike cDNA library"  
/tissue\_type="Spike before anthesis"  
/dev\_stage="Adult plant"  
/lab\_host="E. coli SOLR"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: ECOLI; Site 2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phagescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."  
110 a 145 c 124 g 93 t

## BASE COUNT

Query Match 70.8%; Score 290.2; DB 36; Length 472;  
Best Local Similarity 87.8%; Pred. No. 5.9e-66;  
Matches 316; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 11 cgcggtcatggggcagacaccttgcctcaacattgcagagaagggttccccatctctgt 70  
|||||  
Db 112 GCGCGTCATGGGCGCAGACCTTGCCTCAACATTGCAGAGAAAGGCTTCCCCATCTCTGT 171  
|||||  
QY 71 gtacaacaggaacaacctccaagtggaagcagccgtgcagcgtgcgaaggcagaaggaac 130  
|||||  
Db 172 CTACAACAGGAGCACCCTCCCAAGGTGATGAGACTGTCCAGCGCGCAAGCTAGAGGAAA 231  
|||||  
QY 131 cttcccgctacggtctccatgaccccgctccttcttgagtcattccagaagccacg 190  
|||||  
Db 232 CTTTCCCTGTAGGGTTTCCATGACCTTCATCTCGTCAACTCCATTCAGAACCCACG 291  
|||||  
QY 191 ggtggtgatcatgctgtcagagccgagcgcgcagctgaccagaccatcgagcagctgc 250  
|||||  
Db 292 CGTGTGTCATGCTGTGTCAAGGCTGGCGCACCCAGTTGACACAGACCATCGCCACCTCGC 351  
|||||  
QY 251 agctcacttggagcagggcgactgcatcatcgatggggggaacgagtggtgacgagaacac 310  
|||||  
Db 352 AGCACACCTGGAGCAGGGCGGACTGTCATCGTTGATGAGGAAACGAGTGTATGAGAACAC 411  
|||||  
QY 311 ggag 370  
|||||  
Db 412 AGAAGAGGAGAGAGAGGCGCATGGAGGCGGTGACTCTCTACCTCGGATGGGTGTTTC 471  
|||||

## RESULT 11

BE497880 474 bp mRNA EST 04-AUG-2000  
LOCUS  
DEFINITION WHE0957\_B09\_D172S Wheat pre-anthesis spike cDNA library Triticum aestivum CDNA clone WHE0957\_B09\_D17, mRNA sequence.  
ACCESSION BE497880  
VERSION BE497880.1 GI:9696497  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.  
REFERENCE 1 (bases 1 to 474)  
AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.

**TITLE**  
The structure and function of the expressed portion of the wheat  
genomes - Pre-anthesis spike cDNA library

**JOURNAL**  
Unpublished (2000)

**COMMENT**  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Email: oandersn@pw.usda.gov  
Fax: 5105595818  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.

**FEATURES**  
source  
1. .474  
Location/Qualifiers  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE0957\_B09\_D17"  
/clone\_lib="Wheat pre-anthesis spike cDNA library"  
/tissue\_type="Spike before anthesis"  
/dev\_stage="Adult plant"  
/lab\_host="E. coli SOLR"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site\_1: EcoRI; Site\_2: XhoI; Plants were grown in the  
greenhouse. Whole spike with awns trimmed, white, green  
and yellow anther were collected and total RNA, and  
poly(A) RNA were prepared, a cDNA library was made, and  
the cDNA clones were in vivo excised to give pBluescript  
phagemids in the TJ Close lab (Choi, Close, Fenton) at  
the University of California, Riverside. Plasmid DNA  
preparations and DNA sequencing were performed in the OD  
Anderson lab (all other authors)."

**BASE COUNT** 107 a 143 c 130 g 94 t  
**ORIGIN**

Query Match 70.4%; Score 288.6; DB 36; Length 474;  
Best Local Similarity 87.5%; Pred. No. 1.5e-65;  
Matches 315; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 11 cgcggtcatggtggcagaaccttgcctcaacattgcagagaaagggttccccatctctgt 70  
DB 115 CGCCGTCATGGGCGAGAACCTCGCCCTCAACATTCGGGAGAAAGGTTCCCATCTCTGT 174  
QY 71 gtacacaggacacacctccaagggtgacgagacccgtgcagcgtgccaaggcagaagaaa 130  
DB 175 CTACACAGGAGGACCACTCCCAAGGTGATGAGACTGTTACGCGGCCCAAGCTAGAGGAAA 234  
QY 131 ccttcctctacggtcttccatgaccccgcgctctttgtgaagtcattcagaagccacg 190  
DB 235 CTTTCTCTTACGGTTTCCATGACCCCTGCTATCTCTGCTCAACTCCATTCAAGAGCCACG 294  
QY 191 gdtgtgatcatgctcgtcaaggcggcgccagtcagtcagtcagtcagtcagtcagtcagtc 250  
DB 295 TGTGTCATCATGCTGTGTCAAGGCGGTTGCCCGGTTGCCAGACCAATTGCAAGCGTCGC 354  
QY 251 agctcaattggagcaggcgactgcattgcattgcattgcattgcattgcattgcattgcattgc 310  
DB 355 AGCACACCTGGAGGAGGCGACTGCTGCTGATGAGGAGGACGAGTGGTATGAGAACAC 414  
QY 311 ggagaggaggagagggcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 370  
DB 415 GGAAGGAGGAGAGGCGATGAGGAGGCGTGGACTCTCTACCTCGGGATGGTGTTC 474

**RESULT 12**  
BE425231 512 bp mRNA EST 24-JUL-2000  
**LOCUS** WHE0312\_G04\_G042S Wheat unstressed seedling shoot cDNA library  
**DEFINITION** Triticum aestivum cDNA clone WHE0312\_G04\_G04, mRNA sequence.  
**ACCESSION** BE425231  
**VERSION** BE425231.1 GI:9423146

**KEYWORDS**  
SOURCE ORGANISM  
Triticum aestivum  
Unpublished (2000)

**REFERENCE**  
AUTHORS  
P.S., Hsia,C.C., Kang,Y., Choi,D.W., Close,T.J., Fenton,R.D., Han  
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han  
Seaton,C.B. and Tong,J.C.  
The structure and function of the expressed portion of the wheat  
genomes

**TITLE**  
Unpublished (2000)

**JOURNAL**  
COMMENT  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oandersn@pw.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.

**FEATURES**  
source  
1. .512  
Location/Qualifiers  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE0312\_G04\_G04"  
/clone\_lib="Wheat unstressed seedling shoot cDNA library"  
/tissue\_type="Etiolated shoot"  
/dev\_stage="Five day old seedling"  
/lab\_host="E. coli SOLR"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site\_1: EcoRI; Site\_2: XhoI; Seeds were surface-sterilized  
, germinated and grown aseptically in the dark at room  
temperature on filter paper with water, nystatin and  
cefotaxime in covered crystallization dishes. Shoots were  
harvested. The tissue, total RNA, and poly(A) RNA clones  
prepared, a cDNA library was made, and the cDNA clones  
were in vivo excised to give pBluescript phagemids in the  
TJ Close lab (Choi, Close, Fenton) at the University of  
California, Riverside. Plasmid DNA preparations and DNA  
sequencing were performed in the OD Anderson lab (all  
other authors)."

**BASE COUNT** 116 a 142 c 144 g 110 t  
**ORIGIN**

Query Match 70.4%; Score 288.6; DB 35; Length 512;  
Best Local Similarity 85.8%; Pred. No. 1.6e-65;  
Matches 343; Conservative 0; Mismatches 55; Indels 2; Gaps 2;

QY 11 cgcggtcattggggcagaaccttgcctcaacattgcagagaaagggttccccatctctgt 70  
DB 15 CGCGGTGATGGGCGAGAACCTTGCCCTC-ACATTCGCCGAGAAA-GCTTCCCATCTCTGT 72  
QY 71 gtacacaggacacacctccaagggtgacgagacccgtgcagtcagtcagtcagtcagtcagtc 130  
DB 73 CTACACAGGAGACCACTCCCAAGGTGATGAGACTGTTACGCGTGCACGCTAGAGGAAA 132  
QY 131 ccttcctctacggtcttccatgaccccgcgctctttgtgaagtcattcagaagccacg 190  
DB 133 CTTTCTCTTATGGTTTCCATGACCCCTGCATCTTGTGTAACCTCCATTCAGAACGACG 192  
QY 191 gdtgtgatcatgctcgtcaaggcggcgccagtcagtcagtcagtcagtcagtcagtcagtc 250  
DB 193 TGTGTCATCATGCTGTGTCAAGGCTGCTGCTCGGTCGACGACGACCATCGACGCTCGC 252  
QY 251 agctcaattggagcaggcgactgcattgcattgcattgcattgcattgcattgcattgcattgc 310  
DB 253 AGCACACCTGGAGGAGGCGACTGCTGCTGATGAGGAGGACGAGTGGTATGAGAACAC 312  
QY 311 ggagaggaggagagggcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 370



```

Qy 139 tctacgcttcacgaccccgctcctttgtgaagtcattcagaaagccacggtgtga 198
Db 224 TCTACGCTTCCATGACCCCTGCACTTCCTGCTCAACTCCATTCAGAGCCACGTCGTCA 283
Qy 199 tcatgctcgtcaagggcgccaggtgacgagaccatcgcgagctgcagctcact 258
Db 284 TCATGCTGTCAAGGGCGGTGCCCGTTGACCAGACCACTTCAACGCTCGCAGCACACC 343
Qy 259 tggagcaggcgactgcatcatcgatggggggaacagagtggtgacgagacacagagga 318
Db 344 TGGAGCAGGGCGACTGTCATCGTTGTGAGGGAACGAGTGTATGAGAACACACGGAAGGA 403
Qy 319 ggaagaagccatggagagcgcgcctnctgtatcttgcctggtgtctctggagga 378
Db 404 GGGAGAAGGCGATGGAGAGCGGTGGACTCTCTACCTCGGGATGGGTGTTTCCGGAGGAG 463
Qy 379 agagggtgcccgaacggcccgctccttgatg 410
Db 464 AGGAGGTGCCCGCCATGGCCCATCCATGATG 495

RESULT 15
AW286879
LOCUS LG1_222_F04_b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
DEFINITION
ACCSSION AW286879
VERSION AW286879.2 GI:6858816
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
REFERENCE
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6676723.
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 341
POLYA-No. Location/Qualifiers
1. .463
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
; the library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 109 a 125 c 135 g 94 t
ORIGIN

Query Match 65.6%; Score 269; DB 21; Length 463;
Best Local Similarity 91.9%; Pred. No. 2.le-60;
Matches 284; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 12 ccggtcatggggcagaaaccttgcctcaacattgcagagaagggttccccatctctgtg 71
Db 155 GCGGTCATGGGGCAGAACTTGCCCTCAACATTCGAGAGAAGGCTTCCCATCTCTGTC 214

```

```

Qy 72 tacaacaggacaacctccaaggtgacgagaccggtgacgctgccaagcagaaagaaac 131
Db 215 TACAACAGGACACACCTCCCAAGGTTGATGAGACCGGTGACGCGTCCAGGCGAGAGGAAC 274
Qy 132 cttcccgcttacggcttccatgaccccgctcctttgtgaagttccattcagaagccacg 191
Db 275 CTTCTCTGTGTACGGCTTCATGACCCCTGCTCATCTTGTGAACCTCCATTCAGAAGCCCGGT 334
Qy 192 gtgggtgacatgctcgtcaaggccgagcgccagttgacagaccatcgcgacgctcga 251
Db 335 GTTGTTCATCATGCTTGTCAAGGCTGGTGCACAGTTGACCAGACCATTTGCCGACGCTCGCA 394
Qy 252 gctcacttggagcagcgagcactgcatcatcgatggggggaacgagtgatgacgagaacacg 311
Db 395 GCTCACTTGGAGCAGGTTGACTGTATCATTTGATGAGGGAATGAGTGTATGAGACACT 454
Qy 312 gagaggagg 320
Db 455 GAGAGGAGG 463

```

Search completed: November 4, 2000, 11:52:05  
Job time: 18242 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:06 ; Search time 4352.3 Seconds  
(without alignments)  
448.904 Million cell updates/sec

Title: US-09-300-482-225

Perfect score: 316

Sequence: 1 gataaggtgcgcacactgag.....gaatggaggaggagtactctga 316

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*\*  
1: gb\_est1:\*\*  
2: gb\_est2:\*\*  
3: gb\_est3:\*\*  
4: gb\_est4:\*\*  
5: gb\_est5:\*\*  
6: gb\_est6:\*\*  
7: gb\_est7:\*\*  
8: gb\_est8:\*\*  
9: gb\_est9:\*\*  
10: gb\_est10:\*\*  
11: gb\_est11:\*\*  
12: gb\_est12:\*\*  
13: gb\_est13:\*\*  
14: gb\_est14:\*\*  
15: gb\_est15:\*\*  
16: gb\_est16:\*\*  
17: gb\_est17:\*\*  
18: gb\_est18:\*\*  
19: gb\_est19:\*\*  
20: gb\_est20:\*\*  
21: gb\_est21:\*\*  
22: gb\_est22:\*\*  
23: gb\_est23:\*\*  
24: gb\_est24:\*\*  
25: gb\_est25:\*\*  
26: gb\_est26:\*\*  
27: gb\_est27:\*\*  
28: gb\_est28:\*\*  
29: gb\_est29:\*\*  
30: gb\_est30:\*\*  
31: gb\_est31:\*\*  
32: gb\_est32:\*\*  
33: gb\_est33:\*\*  
34: gb\_est34:\*\*  
35: gb\_est35:\*\*  
36: gb\_est36:\*\*  
37: gb\_est37:\*\*  
38: gb\_est38:\*\*  
39: gb\_est39:\*\*  
40: gb\_est40:\*\*  
41: em\_estba:\*\*  
42: em\_estfun:\*\*  
43: em\_esthum1:\*\*

44: em\_esthum2:\*\*  
45: em\_esthum3:\*\*  
46: em\_esthum4:\*\*  
47: em\_esthum5:\*\*  
48: em\_esthum6:\*\*  
49: em\_esthum7:\*\*  
50: em\_esthum8:\*\*  
51: em\_esthum9:\*\*  
52: em\_esthum10:\*\*  
53: em\_esthum11:\*\*  
54: em\_esthum12:\*\*  
55: em\_esthum13:\*\*  
56: em\_esthum14:\*\*  
57: em\_esthum15:\*\*  
58: em\_esthum16:\*\*  
59: em\_esthum17:\*\*  
60: em\_esthum18:\*\*  
61: em\_esthum19:\*\*  
62: em\_esthum20:\*\*  
63: em\_estin1:\*\*  
64: em\_estin2:\*\*  
65: em\_estin3:\*\*  
66: em\_estin4:\*\*  
67: em\_estom:\*\*  
68: em\_estov1:\*\*  
69: em\_estov2:\*\*  
70: em\_estpl1:\*\*  
71: em\_estpl2:\*\*  
72: em\_estpl3:\*\*  
73: em\_estpl4:\*\*  
74: em\_estpl5:\*\*  
75: em\_estrol:\*\*  
76: em\_estrol2:\*\*  
77: em\_estrol3:\*\*  
78: em\_estrol4:\*\*  
79: em\_estrol5:\*\*  
80: em\_estrol6:\*\*  
81: em\_estrol7:\*\*  
82: em\_estrol8:\*\*  
83: em\_estrol9:\*\*  
84: em\_estrol10:\*\*  
85: em\_estrol11:\*\*  
86: em\_estrol12:\*\*  
87: em\_estrol13:\*\*  
88: gb\_gss1:\*\*  
89: gb\_gss2:\*\*  
90: gb\_gss3:\*\*  
91: gb\_gss4:\*\*  
92: em\_gss1:\*\*  
93: em\_gss2:\*\*  
94: em\_gss3:\*\*  
95: em\_gss4:\*\*  
96: gb\_gss5:\*\*  
97: gb\_gss6:\*\*  
98: gb\_gss7:\*\*  
99: gb\_gss8:\*\*  
100: gb\_gss9:\*\*  
101: em\_gss5:\*\*  
102: em\_gss6:\*\*  
103: em\_gss7:\*\*  
104: em\_gss8:\*\*  
105: em\_gss9:\*\*  
106: em\_gss10:\*\*  
107: em\_gss11:\*\*  
108: gb\_gss10:\*\*  
109: gb\_gss11:\*\*  
110: em\_gss12:\*\*  
111: gb\_gss12:\*\*  
112: gb\_gss13:\*\*  
113: gb\_gss14:\*\*  
114: gb\_gss15:\*\*  
115: gb\_gss16:\*\*  
116: gb\_gss17:\*\*

LOCUS	AU088734	575 bp	mRNA	EST	31-MAR-2000
DEFINITION	AU088734 Rice shoot Oryza sativa cDNA clone s4976, mRNA sequence.				
ACCESSION	AU088734				
VERSION	AU088734.1 GI:7378463				
KEYWORDS	EST.				
SOURCE	Oryza sativa.				
ORGANISM	Oryza sativa				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.				
AUTHORS	1 (bases 1 to 575)				
TITLE	Sasaki, T. and Yamamoto, K.				
JOURNAL	Rice cDNA from etiolated shoot (2000)				
COMMENT	Unpublished (2000)				
	Contact: Takuji Sasaki				
	National Institute of Agrobiological Resources				
	Rice Genome Research Program				
	2-1-2 Kannondai, Tsukuba				
	Ibaraki,				
	Japan 305				
	Tel: 0298-38-7441				
	Fax: 0298-38-7468				
	Email: tsasaki@agr.affrc.go.jp				
	PROJECT = 'RGP'.				
	S4976-8Z.				

```

FEATURES
source
1. 575
Location/Qualifiers
/organism="Oryza sativa"
/strain="Nipponbare, sub_species Japonica."
/db_xref="taxon:4530"
/clone="S4976"
/clone_lib="Rice shoot"
/note="Etiolated shoot (8 days old)"
153 a 109 c 153 g 160 t
PAGE COUNT

```

BASE COUNT	153 a	109 c	153 g	160 t
ORIGIN				
Query Match		50.9%	Score 160.8;	DB 15; Length 575;
Best Local Similarity		83.2%;	Pred. No. 1.3e-39;	
Matches 183;		Conservative	0; Mismatches 37;	Indels 0; Gaps 0;

Qy	1	gataaggtgcgcacactgagaaagaagtcaccttcccttgacatagaggttgatgaggt	60
Db	148	gaaaaggctgggtcccttgnagaagaagtaccatcccttgacatcgagggttgatgctggt	207
Qy	61	ctaggctcttcaacatagacgtggccgcacatctgtgggggccaattgcctgcgtcgga	120
Db	208	ctgggtcccttcaccattgatgtgctgcacatctgctgggctaacTCATggtTCGCTGGA	267
Qy	121	agctctatattggcgtgcgcgaccggagccatcatatctgtctgaggaagaagcgctc	180
Db	268	agctcaaatattttggagacgctcaacccaggagaggtTCATTCGGCACTAAGGAAGCGGT	327

	181	gagggtctcagaacaaaaactgattttgggttcttgct	220
Qy			
	328	GAGGATCTCAGAACAAAAGCTGATCTGGTTTGTACT	367
Dδ			

RESULT	2	EST	30-OCT-1998
LOCUS	AU031584	369 bp	mRNA
DEFINITION	AU031584	Rice cDNA from immature leaf including apical meristem	
		Orza sativa cDNA clone E61890.22, mRNA sequence.	

ACCESSION AU031584  
 VERSION AU031584.1 GI:3767474  
 KEYWORDS  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 369)  
Magnoliophyta; Liliopsida; Poales; Poaceae; *Oryza*.  
Sasaki, T. and Yamamoto, K.  
Rice cDNA from immature leaf including apical meristem  
Unpublished (1997)

RESULT 1  
-- AH088734

1

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160.8	50.9	575	15	AU088734 AU088734 AU088734 AU088734
2	158.8	50.3	369	15	AU031584 AU031584 AU031584 AU031584
3	158.8	50.3	403	36	D22889 RICCI1720A R D22889 RICCI1720A R
4	150.2	47.5	499	36	C98335 C98335 C98335 C98335
5	124.4	39.4	472	36	C73011 C73011 C73011 C73011
6	118.4	37.5	504	14	AL380932 AL380932 AL380932 AL380932
7	118.4	37.5	542	14	AL380933 AL380933 AL380933 AL380933
8	116	36.7	638	21	AW350186 AW350186 AW350186 AW350186
9	116	36.7	736	21	AW348508 AW348508 AW348508 AW348508
10	106.6	33.7	165	35	BE359923 BE359923 BE359923 BE359923
11	100.8	31.9	749	33	BE034126 BE034126 BE034126 BE034126
12	97.2	30.8	483	14	AI993279 AI993279 AI993279 AI993279
13	95	30.1	544	19	AV545174 AV545174 AV545174 AV545174
14	94.6	29.9	619	14	AI999119 AI999119 AI999119 AI999119
15	90.2	28.5	317	39	T75815 T75815 T75815 T75815
16	85.8	27.2	257	21	AW289686 AW289686 AW289686 AW289686
17	85.2	27.0	510	14	AL371121 AL371121 AL371121 AL371121
18	77.8	24.6	653	13	AW896020 AW896020 AW896020 AW896020
19	77.2	24.5	635	21	AW288562 AW288562 AW288562 AW288562
20	70.6	22.3	556	12	AI745746 AI745746 AI745746 AI745746
21	68.2	21.6	546	14	AI947960 AI947960 AI947960 AI947960
22	67.2	21.3	605	11	AI622577 AI622577 AI622577 AI622577
23	66.6	21.1	562	12	AI711616 AI711616 AI711616 AI711616
24	65.4	20.7	525	111	AQ844369 AQ844369 AQ844369 AQ844369
25	63.8	20.2	246	25	AW927596 AW927596 AW927596 AW927596
26	63.8	20.2	475	25	AW928010 AW928010 AW928010 AW928010
27	63	19.9	485	19	AW017710 AW017710 AW017710 AW017710
28	63	19.9	523	19	AW065819 AW065819 AW065819 AW065819
29	63	19.9	535	13	AI783104 AI783104 AI783104 AI783104
30	63	19.9	565	13	AI783103 AI783103 AI783103 AI783103
31	63	19.9	593	19	AW000034 AW000034 AW000034 AW000034
32	62	19.2	562	33	BE056891 BE056891 BE056891 BE056891
33	60.8	19.2	378	14	AI943885 AI943885 AI943885 AI943885
34	60.8	19.2	414	22	AW498391 AW498391 AW498391 AW498391
35	60.8	19.2	460	19	AW052901 AW052901 AW052901 AW052901
36	60.8	19.2	490	36	BE518996 BE518996 BE518996 BE518996
37	59.8	18.9	515	22	AW438351 AW438351 AW438351 AW438351
38	58.2	18.4	495	19	AW091086 AW091086 AW091086 AW091086
39	58	18.4	825	25	AW983286 AW983286 AW983286 AW983286
40	57.8	18.3	424	40	W80360 W80360 W80360 W80360
41	56.8	18.0	448	13	AI168746 AI168746 AI168746 AI168746
42	56.4	17.8	325	9	AT168912 AT168912 AT168912 AT168912
43	55.6	17.6	490	20	AW191777 AW191777 AW191777 AW191777
44	55	17.4	394	14	AI947921 AI947921 AI947921 AI947921
45	55	17.4	613	5	AA556420 AA556420 AA556420 AA556420

## ALIGNMENTS

ORGANISM  
Oryza sativa  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta  
Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
1 (bases 1 to 369)  
Sasaki, T. and Yamamoto, K.  
Rice cDNA from immature leaf including apical meristem  
Unpublished (1997)  
JOURNAL











```

POLYA=Yes.
FEATURES
  source
    Location/Qualifiers
      1..165
        /organism="Sorghum bicolor"
        /db_xref="taxon:4558"
        /clone_lib="Dark Grown 1 (DG1)"
        /note="Organ: 5-day-old dark-grown seedlings; Vector:
        Lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was
        made from poly-A RNA in the cloning vector lambda ZAP II.
        Clones to be sequenced were prepared by mass excision."
BASE COUNT      41 a      46 g      45 t
ORIGIN
Query Match      33.7%; Score 106.6; DB 35; Length 165;
Best Local Similarity 96.5%; Pred. No. 7.9e-23;
Matches 109; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gataaggtgcacactgagaagaagtagacccttccttgacatagaggttgatggtggt 60
|||||
Db 53 GATAAGGTGCGCACATTGAGAAAGAGTACCCTTCCTTGACATAGAGGTTGATGGTGGT 112
|||||

Qy 61 ctaggctctcaaccatagacgtgcccgcacatctgctggggccaaattgcacgt 113
|||||
Db 113 TTAGGCCCTTCAACCATAGACGTGTCCGCATCTGCTGGGCCAATTGTCATCGT 165
|||||

RESULT 11
LOCUS      BE034126      749 bp      mRNA      EST      07-JUN-2000
DEFINITION MG05C09 MG Mesembryanthemum crystallinum cDNA 5' similar to
d-ribulose-5-phosphate 3-epimerase, mRNA sequence.
ACCESSION      BE034126
VERSION      BE034126.1 GI:8329120
KEYWORDS      EST.
SOURCE      common ice plant.
ORGANISM      Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE      Bohnert, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., Ferrea
AUTHORS      H., Kawasaki, S., McColough, A., Michalowski, C.B., Palacio, C.,
Scara, G., Wheeler, M. and Zepeda, G.R.
TITLE      Functional Genomics of Plant Stress Tolerance
JOURNAL      Unpublished (2000)
COMMENT      Contact: Michalowski, C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu
An open reading frame exists.
FEATURES
  source
    Location/Qualifiers
      1..749
        /organism="Mesembryanthemum crystallinum"
        /db_xref="taxon:3544"
        /clone_lib="MG"
        /tissue_type="roots"
        /dev_stage="5-6 weeks"
        /note="grown in hydroponics, stress 400 mM NaCl (in 0.5
        Hoagland's), 30 h stress"
BASE COUNT      207 a      194 g      213 t      2 others
ORIGIN
Query Match      31.9%; Score 100.8; DB 33; Length 749;
Best Local Similarity 69.0%; Pred. No. 7.1e-21;
Matches 138; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 1 gataaggtgcacactgagaagaagtagacccttccttgacatagaggttgatggtggt 60
|||||
Db 327 GATAAGGTTCGTCGTTGAGGAGAGTAGTCCATCTCTTGATATTGAGGTGATGGTGGT 386
|||||

```

```

Qy 61 ctaggctctcaaccatagacgtgcccgcacatctgctggggccaaattgcacgtcgctgga 120
|||||
Db 387 TTGGGGCCATCAACCATTTGATGCTGCGCAAGCTGGAGCAAAATTCGATTTGTTCTGGG 446
|||||

Qy 121 agctctatatattgctgcggaccagagccatcatctatctgctgaggaagagcgctc 180
|||||
Db 447 AGTTCAGTGTGTTGGTCCCTTGAAACAGCTCAAGTATCTCGTTATTGAGGAAGACGGTG 506
|||||

Qy 181 gagggctctcagacacaaaaa 200
|||||
Db 507 GAGCAAGCCCGCAGCGGTAACA 526
|||||

RESULT 12
LOCUS      AI993279      483 bp      mRNA      EST      08-SEP-1999
DEFINITION 701495920 A. thaliana, Ohio State clone set Arabidopsis thaliana
cDNA clone 701495920, mRNA sequence.
ACCESSION      AI993279
VERSION      AI993279.1 GI:5840184
KEYWORDS      EST.
SOURCE      thale cress.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE      1 (bases 1 to 483)
AUTHORS      Chen, J., Horiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R.,
Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.,
Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrigha, A., Murry, L.,
Turner, C., Krikorian, S., Elder, L. and Hanson, D.
TITLE      Arabidopsis thaliana Gene Expression MicroArray
JOURNAL      Unpublished (1999)
COMMENT      Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.
FEATURES
  source
    Location/Qualifiers
      1..483
        /organism="Arabidopsis thaliana"
        /db_xref="taxon:3702"
        /clone="701495920"
        /clone_lib="A. thaliana, Ohio State clone set"
        /note="cDNA library was made from selected clones from the
        Arabidopsis thaliana Ohio State clone set."
BASE COUNT      120 a      109 c      121 g      133 t
ORIGIN
Query Match      30.8%; Score 97.2; DB 14; Length 483;
Best Local Similarity 67.0%; Pred. No. 8.5e-20;
Matches 138; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 1 gataaggtgcacactgagaagaagtagacccttccttgacatagaggttgatggtggt 60
|||||
Db 31 GACAAAGTCCGGGCGATTGAGGCATATAATACCAACTTGATATTCAGTGGATGGCGGC 90
|||||

Qy 61 ctaggctctcaaccatagacgtgcccgcacatctgctggggccaaattgcacgtcgctgga 120
|||||
Db 91 TTAGGCCCTTCAACGATCGTCAAGCGGCTGCAGCGCGCAACTGTATTGTTCTCTGGA 150
|||||

Qy 121 agctctatatattgctgcggaccagagccatcatctgctgaggaagagcgctc 180
|||||
Db 151 AGTTCAGTGTGTTGGAGCTCCCGAGCGAGGATGTCTCTCTCTCTCTCTCTCTCTCTCT 210
|||||

Qy 181 gagggctctcagacacaaaaactgatt 206
|||||

```



Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
1 (bases 1 to 317)  
Newman, T., deRuijn, F. J., Green, P., Keegstra K., Kende, H., McIntosh  
L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomasow, M., Retzel  
E. and Somerville, C.

**TITLE** Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones

**JOURNAL** Plant Physiol. 106, 1241-1255 (1994)

JOURNAL OF  
PHYSICS: 1997  
MEDLINE  
95148729

On Apr 14, 1993 this sequence replaced gi:692577.

MSU-DOE Plant Research Laboratory  
Michigan State University  
Contact: Thomas Newman  
msu-doe-plant-research-lab@msu.edu

MSU=DOE-PRL, MICHAEL  
 TANCING MICHAEL

Lansing, MI  
Tel. 517-353-0954

Tel: 517-353-0854  
Fax: 517-353-0168

Fax: 517-353-9168  
Email: 20213+cc@bmcj.msu.edu

Email: 22313tcn@ibm.ci.msu.edu  
 doc output: m7.doc primer

Seq primer: T7 dye primer.

Location/Qualifiers

1. .317

/organism="Arabidopsis

```
/strain="var colu
```

```
/db_xref="taxon:3702"
```

/clone="146H5T7"  
/clone\_lib="Lambda-PR1.2"  
/note=vector: lambda Zip-Lox; Site\_1: Sal; Site\_2: Not;  
Lambda PR1.2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dt primed cDNA. "

ORIGIN OF PRIMED CBNs:		
BASE COUNT	74 a	68 c
	85 a	85 c
	79 t	79 t
	11 others	11 others

BASE CO  
ORIGIN

[illegible]

Search completed: November 4, 2000, 11:52:10  
Job time: 18247 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:21 ; Search time 4352.3 Seconds  
(without alignments)  
426.175 Million cell updates/sec

Title: US-09-300-482-619

Perfect score: 300

Sequence: 1 gattattgacaacacgat.....ttcagaaatgtctgtgtg 300

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est25:\*
- 26: gb\_est26:\*
- 27: gb\_est27:\*
- 28: gb\_est28:\*
- 29: gb\_est29:\*
- 30: gb\_est30:\*
- 31: gb\_est31:\*
- 32: gb\_est32:\*
- 33: gb\_est33:\*
- 34: gb\_est34:\*
- 35: gb\_est35:\*
- 36: gb\_est36:\*
- 37: gb\_est37:\*
- 38: gb\_est38:\*
- 39: gb\_est39:\*
- 40: gb\_est40:\*
- 41: em\_estba:\*
- 42: em\_estfun:\*
- 43: em\_esthum1:\*

44: em\_esthum2:\*

45: em\_esthum3:\*

46: em\_esthum4:\*

47: em\_esthum5:\*

48: em\_esthum6:\*

49: em\_esthum7:\*

50: em\_esthum8:\*

51: em\_esthum9:\*

52: em\_esthum10:\*

53: em\_esthum11:\*

54: em\_esthum12:\*

55: em\_esthum13:\*

56: em\_esthum14:\*

57: em\_esthum15:\*

58: em\_esthum16:\*

59: em\_esthum17:\*

60: em\_esthum18:\*

61: em\_esthum19:\*

62: em\_esthum20:\*

63: em\_estin1:\*

64: em\_estin2:\*

65: em\_estin3:\*

66: em\_estin4:\*

67: em\_estom:\*

68: em\_estovi:\*

69: em\_estov2:\*

70: em\_estpl1:\*

71: em\_estpl2:\*

72: em\_estpl3:\*

73: em\_estpl4:\*

74: em\_estpl5:\*

75: em\_estro1:\*

76: em\_estro2:\*

77: em\_estro3:\*

78: em\_estro4:\*

79: em\_estro5:\*

80: em\_estro6:\*

81: em\_estro7:\*

82: em\_estro8:\*

83: em\_estro9:\*

84: em\_estro10:\*

85: em\_estro11:\*

86: em\_estro12:\*

87: em\_estro13:\*

88: gb\_gss1:\*

89: gb\_gss2:\*

90: gb\_gss3:\*

91: gb\_gss4:\*

92: em\_gss1:\*

93: em\_gss2:\*

94: em\_gss3:\*

95: em\_gss4:\*

96: gb\_gss5:\*

97: gb\_gss6:\*

98: gb\_gss7:\*

99: gb\_gss8:\*

100: gb\_gss9:\*

101: em\_gss5:\*

102: em\_gss6:\*

103: em\_gss7:\*

104: em\_gss8:\*

105: em\_gss9:\*

106: em\_gss10:\*

107: em\_gss11:\*

108: gb\_gss10:\*

109: gb\_gss11:\*

110: em\_gss12:\*

111: gb\_gss12:\*

112: gb\_gss13:\*

113: gb\_gss14:\*

114: gb\_gss15:\*

115: gb\_gss16:\*

116: gb\_gss17:\*

117: gb\_gss18.\*  
 118: gb\_gss19.\*  
 119: em\_gss13.\*  
 120: gb\_gss20.\*  
 121: gb\_gss21.\*  
 122: gb\_gss22.\*  
 123: gb\_gss23.\*  
 124: gb\_gss24.\*  
 125: em\_gss14.\*  
 126: em\_gss15.\*  
 127: em\_gss16.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	285.6	95.2	573	12	AI657311	AI657311 486093C03
2	225.2	75.1	423	23	AW564778	AW564778 LG1_301.C
3	223.8	74.6	330	36	D43256	D43256 D43256 Rice
4	210.4	70.1	550	34	BE205297	BE205297 EST397973
5	208.8	69.6	419	19	AW035534	AW035534 EST281272
6	191.4	63.8	670	23	AW584861	AW584861 N211008e
7	182.4	60.8	538	25	AW933758	AW933758 EST359601
8	172	57.3	537	22	AW399786	AW399786 EST310286
9	135.8	45.3	597	35	BE443964	BE443964 WHE1123_E
10	133	44.3	217	14	AL366975	AL366975 MCB1123_E
11	106.6	35.5	501	21	AW254881	AW254881 ML1306 pe
12	104	34.7	621	21	AW255839	AW255839 ML923 pep
13	94.6	31.5	564	11	AI482995	AI482995 EST242318
14	89	29.7	439	35	BE400087	BE400087 AWB010.F1
15	83	27.9	588	18	AW522495	AW522495 AV522495
16	71.6	23.9	681	21	AW256424	AW256424 EST304491
17	71	23.7	452	113	AQ953823	AQ953823 nbe0053A
18	61	20.3	404	37	L33519	L33519 BNAESTF092
19	58.6	19.5	326	19	AV551048	AV551048 AV551048
20	42.8	14.3	465	114	AZ046382	AZ046382 nbe00085N
21	39.8	13.3	533	23	AW569848	AW569848 s182802.Y
22	37.8	12.6	506	90	AQ240320	AQ240320 CIT-HSP-2
23	34	11.3	240	15	AU074940	AU074940 AU074940
24	34	11.3	704	15	AU061833	AU061833 AU061833
25	33.6	11.2	254	22	AW484099	AW484099 59215 MAR
26	33.6	11.2	1045	124	CNS005JIG	AL340225 Tetraodon
27	33.4	11.1	266	36	BE529610	BE529610 M75E07STM
28	33.2	11.1	416	37	F15380	F15380 ATT55628 G1
29	33.2	11.1	546	9	A1260448	A1260448 LP04319.5
30	32.6	10.9	946	121	CNS006RN	AL065838 Drosophila
31	32.4	10.8	369	22	AW430628	AW430628 70368 MAR
32	32.2	10.7	458	117	AZ281677	AZ281677 RPCI-23-1
33	32.2	10.7	611	114	AZ064575	AZ064575 RPCI-23-4
34	32	10.7	191	29	BB273748	BB273748 BB273748
35	32	10.7	851	131	CNS00AEL	AL055177 Drosophila
36	31.6	10.5	334	9	A1217424	A1217424 qb22h09.x
37	31.6	10.5	925	121	CNS00LEK	AL068320 Drosophila
38	31.4	10.5	369	3	AA327905	AA327905 EST31307
39	31.4	10.5	431	8	A1034194	A1034194 ox12h07.x
40	31.4	10.5	451	10	AI452153	AI452153 mp80h08.x
41	31.4	10.5	468	20	AW221197	AW221197 EST297666
42	31.2	10.4	372	33	BE109286	BE109286 UI-R-BJ1-
43	31.2	10.4	399	22	AW531748	AW531748 UI-R-C4-a
44	31.2	10.4	457	1	AA004451	AA004451 zh89c05.s
45	31.2	10.4	492	32	BE020509	BE020509 sm44f04.y

## ALIGNMENTS

RESULT 1  
 AI657311

LOCUS AI657311 573 bp mRNA EST 05-MAY-1999  
 DEFINITION 486093C03.y1 486 - leaf primordia cDNA library from Hake lab Zea mays cDNA, mRNA sequence.  
 ACCESSION AI657311  
 VERSION AI657311.1 GI:4753401  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 REFERENCE Walbot.V.  
 AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford University  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Walbot V  
 COMMENT Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 486093 row: C column: 03.  
 FEATURES  
 source  
 1..573  
 /location="Qualifiers  
 /organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="486 - leaf primordia cDNA library from Hake lab"  
 /tissue\_type="leaf primordia"  
 /dev\_stage="p7-p11 leaf"  
 /db\_host="E.coli XL1-Blue MFR"  
 /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA library."  
 BASE COUNT 161 a 115 c 133 g 164 t  
 ORIGIN  
 Query Match 95.2%; Score 285.6; DB 12; Length 573;  
 Best Local Similarity 97.0%; Pred. No. 41e-79;  
 Matches 291; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 gatttattgcaacacacgacccctgctgggattgattcatcaataattgctcaactagagacctg 60  
 |||||  
 DB 61 GATTTATTGACACACCGATCTCTGGGATTGATCATCAAAATTGCTCAACTAGGACCTG 120  
 |||||  
 QY 61 aactggcaactactcttctgttaattgtcatttcttaagcgaggagcacactgaaacccgca 120  
 |||||  
 DB 121 AACTGGCACTACTCTTGTAAATTGTCATTTCTTAAGAGCGGAGCACACCTGAAACCCGCA 180  
 |||||  
 QY 121 atggtctactagaagtacagaagaccttcagagatcggggctgcaattctcgaacacagg 180  
 |||||  
 DB 181 ATGGGCTACTGGAGTACAGAAAGCCCTTCAGAGATGCGAGTCTGCAATTTCTCGAAACAGG 240  
 |||||  
 QY 181 gtgttgcattactcaagaaattctctgttgataacactgctagatagagaggtggt 240  
 |||||  
 DB 241 GTGTTGCAATCACTCAAGAAATTTCTGTGTGATAACACTGCCAGATAGAGGGATGTT 300  
 |||||  
 QY 241 tagctcgggttctctatgtttgattgggttggttgtaggacttcagaaatctctgctgtg 300  
 |||||  
 DB 301 TAGTCGGTTTCTTATGTTTACTGGGTTGGTGGTAGGACTTCAGAAATGTCAGCTGTG 360  
 |||||  
 RESULT 2  
 AW564778 423 bp mRNA EST 19-JUL-2000  
 LOCUS LG1\_301\_C06.bl\_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA  
 DEFINITION sequence.  
 ACCESSION AW564778  
 VERSION AW564778.1 GI:7218656  
 KEYWORDS EST.  
 SOURCE sorghum.





Sat Nov 4 18:11:44 2000

http://chrysie.tamu.edu/medicago  
Seq primer: Skmod (CTA gAA CTA gTg gAT CC).

## FEATURES

Location/Qualifiers  
1. 550  
/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="PKV0-21015"  
/clone\_lib="KV0"  
/tissue\_type="Seedling roots"  
/dev\_stage="Immediately prior to inoculation with  
Sinorhizobium meliloti (0 hour)"  
/lab\_host="E. coli strain SOLR"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from polyA+ enriched RNA. The  
cDNA was directionally ligated into the Unizap XR vector  
from Stratagene and packaged using Gigapack III Gold  
packaging extracts. Plasmids containing cDNA inserts  
were excised from the recombinant lambda-Zap phage using  
Ex-assist helper phage and propagated in SOLR cells."  
BASE COUNT 143 a 115 c 139 g 153 t  
ORIGIN

Query Match 70.1%; Score 210.4; DB 34; Length 550;  
Best Local Similarity 81.3%; Pred. No. 1.6e-55;  
Matches 244; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 1 gatttattgacacaccgacatcctgctggcattgacatcaaatgctcaactaggacctg 60  
Db 126 GATTATTGACAAACACAGATCCTGCTGGCATTGATCATCATGATTCGACAACTCGGCCCTG 185  
QY 61 aactggcaactactcttgaattgctatttgaagagcgaggacacacctgaaccgcga 120  
Db 186 AGCTAGCTTCGACACTTGAATTTGATTTCAAGAGTGGAGGACCCCTCGAGACTAGAA 245  
QY 121 atggtctactagaagtagacagaagccttcagagatgcgggctgcaattctcgaacagg 180  
Db 246 ATGGCTTGTAGAGTACAAAGGCTTTCGTGAGCAGGATGGATTTCCTCCAAACAGG 305  
QY 181 gtgttgcaattactcaagaaattctctgttggaataacactgctagatagagggatgg 240  
Db 306 GTGTGCTATAACACAGAAACTCTTGTGGATAAACACTGCCAGAAATTGAGGGATGGT 365  
QY 241 tagctcggttctactgttattgattgggttggtgtagacttcagaaatgctctgctgg 300  
Db 366 TGGCTAGATTTCCTCCATGTTGACTGGGTGGAGGTAGAACGTCAGAGATGCTGCAGTCG 425

RESULT 5  
AW035534 419 bp mRNA EST 15-SEP-1999  
LOCUS  
DEFINITION EST281272 tomato callus, TAMU Lycopersicon esculentum cDNA clone  
cLEC39G17 similar to glucose-6-phosphate isomerase, putative, mRNA  
sequence.  
ACCESSION AW035534  
VERSION AW035534.1 GI:5894290  
KEYWORDS EST.  
SOURCE Tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids  
I; Solanales; Solanales; Solanaceae; Solanum; Lycopersicon.  
REFERENCE 1 (bases 1 to 419)  
AUTHORS Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,  
Liang, F., Opton, J., Craven, M.B., Bowman, C.L., Ahn, S., Rønning,  
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.  
Generation of ESTs from tomato callus tissue  
Unpublished (1999)  
CONTACT: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU  
5 prime sequence.  
Location/Qualifiers  
1. 419  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEC39G17"  
/clone\_lib="tomato callus, TAMU"  
/tissue\_type="callus"  
/dev\_stage="25-40 days old"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons  
of seedlings 7-10 days post-germination were excised, cut  
at both ends and placed on MS medium with no selection.  
Mixed callus was harvested at 25 and 40 days and included  
undifferentiated masses. Tomato Callus EST Library"

BASE COUNT 109 a 92 c 102 g 116 t  
ORIGIN

Query Match 69.6%; Score 208.8; DB 19; Length 419;  
Best Local Similarity 81.0%; Pred. No. 4.8e-55;  
Matches 243; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 1 gatttattgacacaccgacatcctgctggcattgacatcaaatgctcaactaggacctg 60  
Db 112 GATTATTGACAAACACAGATCCTGCTGGCATTGATCATCAATTCGCAAACTGGCCCTG 171  
QY 61 aactggcaactactcttgaattgctatttgaagagcgaggacacacctgaaccgcga 120  
Db 172 AGCTGCTTCAACACTTGTATTGTTATTTCAAGAGTGGAGGCACTCCAGAAACCCGAA 231  
QY 121 atggtctactagaagtagacagaagccttcagagatgcgggctgcaattctcgaacagg 180  
Db 232 ATGGCTTGTAGAGTTCAGAGGCTTCCTGGAAGCTGGCCTGATATTTCGCAAAACAGG 291  
QY 181 gtgttgcaattactcaagaaattctctgttggaataacactgctagatagagggatgg 240  
Db 292 GTGTGCTATACACAGAAACTCATTTGCTTGCACAACTGCTAGAAATTGAGGGTGGT 351  
QY 241 tagctcggttctactgttattgattgggttggtgtaggacttcagaaatgctctgctgg 300  
Db 352 TGGCCAGATTCCCTATGTTGATTGGTGGCGGAGAACCTCAGAAATGCTCTGCAGTTG 411

RESULT 6  
AW584861 670 bp mRNA EST 28-APR-2000  
LOCUS  
DEFINITION N211008e MHAM Medicago truncatula/Glomsu versiforme mixed EST  
library cDNA clone MHAM-7H23, mRNA sequence.  
ACCESSION AW584861  
VERSION AW584861.1 GI:7261915  
KEYWORDS EST.  
SOURCE Medicago truncatula/Glomsu versiforme mixed EST library.  
ORGANISM Medicago truncatula/Glomsu versiforme mixed EST library  
Eukaryota; mixed EST libraries.  
REFERENCE 1 (bases 1 to 670)  
AUTHORS Harrison, M.J., Liu, J., Peng, H., Gonzales, M., Ellis, L., Town, C.D.,  
Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.  
ESTs from roots of Medicago truncatula after colonization with  
Glomsu versiforme  
Unpublished (2000)  
CONTACT: Harrison, M.J.  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73401  
Tel: 580-223-5810  
Fax: 580-221-7380  
Email: mjharrison@noble.org





**TITLE**  
Medicago truncatula ESTs from nitrogen-starved roots  
**JOURNAL**  
Unpublished (2000)  
**COMMENT**  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de  
Biologie Moleculaire des Relations Plantes-Microorganismes,  
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :  
MT-est@toulouse.inra.fr Website :  
http://sequence.toulouse.inra.fr/Mtruncatula.html).  
**FEATURES**  
Location/Qualifiers  
1. .217  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/db\_xref="taxon:3880"  
/clone="MCBAILD05"  
/clone\_lib="MtBA"  
/tissue\_type="root tips"  
/dev\_stage="harvested after 3 days of N-starvation"  
/note="Vector: pBluescript PSK; Site.1: EcoRI; Site.2:  
XhoI; Plants were grown in an aeroponic chamber for 14  
days on nitrogen-rich medium followed by 3 days on N-free  
medium. RNA was extracted from root tips (1-3 cm). cDNA  
was prepared from polyA+ enriched RNA. The cDNA was  
directionally ligated into Uni-zapR vector from  
Stratagene and packaged using Gigapack Gold packaging  
extracts. Plasmids containing cDNA inserts were  
mass-excised from phage stocks using EXassit helper phage  
and propagated in SOLR cells. Clone ordering and  
sequencing was performed by the Centre National de  
Sequencage (Genoscope, Evry, France)."  
BASE COUNT 58 a 37 c 60 g 61 t 1 others  
ORIGIN

Query Match 44.3%; Score 133; DB 14; Length 217;  
Best Local Similarity 78.4%; Pred. No. 2.5e-31;  
Matches 171; Conservative 0; Mismatches 46; Indels 1; Gaps 1;  
QY 58 ctgaactggaactactcttgaattgtcttaattcttaagagcggagggacacactgaaccc 117  
Db 1 CTGAGCTAGCTTCGACACTTGTGATTGTGATTTCAAAGAGTGGAGGCCCTTACAGTA 60  
QY 118 gcaatggtctactagaagcagaaagcctcagagatcgcgggctgcaattctcgaac 177  
Db 61 GAATGGCTTGTAGAGT-CAAAAGGCTTCGTGAAGCAGGATTGGATTCCCAAAAC 119  
QY 178 aggtgttgcaattactcaagaaaattctctgttggtataacactgctagaatagggat 237  
Db 120 AGGGTGTCTATTACACAAGAAACTCTTCTTGGATAAACAACCTGCCAGATTGAGGAT 179  
QY 238 ggttagctcggttctctatgttattggttgggtggtggt 275  
Db 180 GGTGGCTAGATTCCCATGTTTGAAGTGGTGGGGAGGT 217

**RESULT** 11  
AW254881 501 bp mRNA EST 15-MAR-2000  
**LOCUS** ML1306 peppermint glandular trichome Mentha x piperita cDNA, mRNA  
**DEFINITION** sequence.  
**ACCESSION** AW254881  
**VERSION** AW254881.1 GI:7244261  
**KEYWORDS** EST.  
**SOURCE** peppermint.  
**ORGANISM** Mentha x piperita  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids  
I; Lamiales; Lamiaceae; Mentha.  
**REFERENCE** 1 (bases 1 to 501)

**AUTHORS**  
Lange,B.M., Wildung,M.R., Stauber,E.J., Sanchez,C., Pouchnik,D. and  
Croteau,R.  
**TITLE**  
Probing essential oil biosynthesis and secretion by functional  
evaluation of expressed sequence tags from mint glandular trichomes  
**JOURNAL**  
Proc. Natl. Acad. Sci. U.S.A. 97 (6), 2934-2939 (2000)  
**COMMENT**  
Contact: Lange, B.M.  
Institute of Biological Chemistry/Washington State University  
Pullman, WA  
Email: lange-m@mail.wsu.edu.  
**FEATURES**  
Location/Qualifiers  
1. .501  
/organism="Mentha x piperita"  
/cultivar="Black Mitcham"  
/db\_xref="taxon:34256"  
/clone\_lib="peppermint glandular trichome"  
/tissue\_type="pellate glandular trichomes"  
/cell\_type="secretory"  
/note="Vector: lambda ZAPII"  
BASE COUNT 145 a 105 c 129 g 122 t  
ORIGIN

Query Match 35.5%; Score 106.6; DB 21; Length 501;  
Best Local Similarity 76.9%; Pred. No. 6.1e-23;  
Matches 130; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 1 gatttattgaaacacgcatctctgttggttgatcatcaaatgtctcaactagacctg 60  
Db 333 GATTTCATTGACATACAGATCCAGCGGATTCATCATCAGATAGCACAACTTGGCAGTG 392  
QY 61 aactggaactactcttgtaattgtctattctaaagcggagggacacactgaacccgca 120  
Db 393 AGCTAATCTACACTCGCTGATGATGTTTCAAGAGTGGAGTACGCCGGAACAGAA 452  
QY 121 atgtctactagaagtacagaagccttcagagatcgcggggtcgcaatt 169  
Db 453 ACGGTCTACTTGAAGTTCAGAAAGCCTTCGAGAAGCTGCTCTGGATT 501

**RESULT** 12  
AW255839 621 bp mRNA EST 15-MAR-2000  
**LOCUS** ML923 peppermint glandular trichome Mentha x piperita cDNA, mRNA  
**DEFINITION** sequence.  
**ACCESSION** AW255839  
**VERSION** AW255839.1 GI:7245091  
**KEYWORDS** EST.  
**SOURCE** peppermint.  
**ORGANISM** Mentha x piperita  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids  
I; Lamiales; Lamiaceae; Mentha.  
**REFERENCE** 1 (bases 1 to 621)  
**AUTHORS** Lange,B.M., Wildung,M.R., Stauber,E.J., Sanchez,C., Pouchnik,D. and  
Croteau,R.  
**TITLE**  
Probing essential oil biosynthesis and secretion by functional  
evaluation of expressed sequence tags from mint glandular trichomes  
**JOURNAL**  
Proc. Natl. Acad. Sci. U.S.A. 97 (6), 2934-2939 (2000)  
**COMMENT**  
Contact: Lange, B.M.  
Institute of Biological Chemistry/Washington State University  
Pullman, WA  
Email: lange-m@mail.wsu.edu.  
**FEATURES**  
Location/Qualifiers  
1. .621  
/organism="Mentha x piperita"  
/cultivar="Black Mitcham"  
/db\_xref="taxon:34256"  
/clone\_lib="peppermint glandular trichome"  
/tissue\_type="pellate glandular trichomes"  
/cell\_type="secretory"  
/note="Vector: lambda ZAPII"  
BASE COUNT 181 a 138 c 156 g 145 t  
ORIGIN

Query Match 34.7%; Score 104; DB 21; Length 621;  
 Best Local Similarity 72.2%; Pred. No. 4.3e-22;  
 Matches 148; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 1 gattattgacacacccgactctgctgggattgatcatcaaatgctcaactaggacctg 60  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 367 GATTATTGACAATACAGATCCAGCCGGATTGATCATCATGACACACTTGGCAGTG 426  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 61 aactggcaactactctgttaattgcaattcttaagaagcggagcacacctgaacccgca 120  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 427 AGCTAGAACTTACACTCGTTATGTTGTTTCAAAAGAGTGAAGTACGCCGAAACAAGA 486  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 121 atggtctactagaagacagcctt-cagagatgcggcgctgcaattctcgaacag 179  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 487 AACGTCTACTTGAAGTTCAAAAAGCCTTCCGAGAGCCGTGCTGATTTCCCAACAG 546  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 180 ggtgttgcattactcaagaaatt 204  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 547 GGTCTGCCAATTCCTCCAAAAAAT 571

RESULT 13  
 AI482995 564 bp mRNA EST 29-JUN-1999  
 LOCUS EST242318 tomato shoot, Cornell Lycopersicon esculentum cDNA clone  
 DEFINITION cLEB3B18 similar to glucose-6-phosphate isomerase, mRNA sequence.  
 ACCESSION AI482995  
 VERSION AI482995.1 GI:4386919  
 KEYWORDS EST  
 SOURCE Tomato.  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids  
 I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 564)  
 van der Hoeven, R.S., Matern, A.L., Vision, T., Holt, I.E., Liang, F.,  
 Upton, J., Konning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L.,  
 Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Giovannoni, J.J.  
 and Tanksley, S.D.

TITLE Generation of ESTs from tomato shoot meristem

JOURNAL Unpublished (1999)

COMMENT Other ESTs: EST242322

Contact: David Frisch

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 4366

Fax: 864 656 4293

Email: dfrisch@CLEMSON.EDU.

Location/Qualifiers

1. .564

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="cLEB3B18"

/clone\_lib="tomato shoot, Cornell"

/tissue\_type="shoot meristem"

/dev\_stage="8 week old plants"

/lab\_host="XLOR"

/note="Vector: pBK\_CMV; Site\_1: EcoRI; Site\_2: XhoI; cLEB

- Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA

library made from tomato vegetative shoots including

meristems and small expanding leaves."

BASE COUNT 140 a 103 c 154 g 167 t

ORIGIN

Query Match 31.58; Score 94.6; DB 11; Length 564;

Best Local Similarity 82.08; Pred. No. 3.7e-19;

Matches 109; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 gattattgacacacccgactctgctgggattgatcatcaaatgctcaactaggacctg 60  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 432 GATTATTGACAATACAGATCCAGCCGGATTGATCATCAATTCGACAACTTGGCCCTG 491  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 61 aactggcaactactctgttaattgcaattcttaagaagcggagcacacctgaacccgca 120  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 492 AGCTGGCTTCAACACTTGTATTGTTTCAAGAGTGGAGGCACTCCAGAAACCCGAA 551  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 121 atggtctactaga 133  
 ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 552 ATGGCTTGCTAGA 564

RESULT 14

BE400087 439 bp mRNA EST 21-JUL-2000

LOCUS AWB010.F10F000328 ITEC AWB Wheat Meiotic Stage Library Triticum

DEFINITION aestivum cDNA clone AWB010.F10, mRNA sequence.

ACCESSION BE400087

VERSION BE400087.1 GI:9359555

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.

REFERENCE 1 (bases 1 to 439)

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier,

S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,

Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudriet, P.,

Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,

Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,

Sorrells, M., Warburton, M. and Wenzel, G.

International Triticace EST Cooperative (ITEC): Production of

Expressed Sequence Tags for Species of the Triticaceae

Unpublished (2000)

Contact: Langridge P

Special Research Centre, Basic and Applied Plant Molecular Biology

Waite Campus, University of Adelaide, Glen Osmond SA 5064 AUSTRALIA

Tel: 61 8 8303 7368

Fax: 61 8 8303 7102

Email: plangridge@waite.adelaide.edu.au

International Triticaceae EST Cooperative (ITEC)

http://wheat.pw.usda.gov/genome.

Location/Qualifiers

1. .439

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

/clone="AWB010.F10"

/tissue\_type="excised florets"

/dev\_stage="meiotic stage no later than metaphase 1"

/note="Vector: pSport 1 (Life Technologies cat. no. 18248-013); Site\_1: SalI; Library constructed in pSport 1.

Directionally cloned using the Superscript Plasmid System

for cDNA synthesis and plasmid cloning. M13 Reverse

sequencing primer used to obtain 5' sequence data. 1.4

Kbp average insert size."

BASE COUNT 135 a 81 c 105 g 118 t

ORIGIN

Query Match 29.7%; Score 89; DB 35; Length 439;

Best Local Similarity 86.7%; Pred. No. 2e-17;

Matches 98; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 188 aattactcaagaaaattctctgttgataacactgtagaggtgtagctcg 247

||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

Db 1 AATCACTCAAGAAAATTCCTCTATTAGATAACACTGCCACATAGAGGATGGTTAGATCG 60

||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

QY 248 gtttcctatgtttgattgggttggtggtgtaggacttcagaaatgtctgtg 300

||||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

Db 61 ATTCTCTATGTTGACTGCTTGGTGGTAGAACATCAGAAATGTTCAGCGGGG 113

||||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

Query Match	27.7%;	Score 83;	DB 18;	Length 588;
Best Local Similarity	77.1%;	Pred. No. 1.6e-15;		
Matches 101;	Conservative 0;	Mismatches 30;	Indels 0;	Gaps 0;
QY 170	ctcgaaacagggtgttgcaattactcaagaaattctgtgttgataaacactgtaaat	229		
Db 1	CGCAAACAGGGTGTGCAATACCCCAAGAAACTCATTACTGGATAATACGCCAGAAT	60		
QY 230	agagggatggttagctcggtttctctatgtttgttgattgggttggttaggacttcagaaat	289		
Db 61	TGAAGGTTCGCTAGCTAGATTTCCATGTACGACTCGGTGGGTGGGAACATCAATAAT	120		
QY 290	gtctgctgtgg	300		
Db 121	GTCTGCAGTTG	131		

Search completed: November 4, 2000, 11:52:25  
Job time: 18262 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:34:08 ; Search time 189.35 Seconds  
(without alignments)  
307.515 Million cell updates/sec

Title: US-09-300-482-356

Perfect score: 385

Sequence: 1 caaccgacacacagcctgac.....acggtctggaggttttttt 385

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_NA.\*  
2: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	170.4	44.3	2629	4	US-09-012-030-1
2	170.4	44.3	2629	4	US-08-590-454-1
3	76.8	19.9	1995	6	PCT-US96-05320A-1025
4	51.6	13.4	1872	4	US-08-743-637B-17
5	51.6	13.4	1872	5	US-08-526-840B-17
6	41.2	10.7	4195	2	US-08-340-011-1
7	41.2	10.7	4195	5	US-08-901-710-1
8	41.2	10.7	4425	1	US-08-222-615-31
9	41.2	10.7	4425	6	PCT-US95-04228-31
10	41.2	10.7	4795	2	US-08-340-011-3
11	41.2	10.7	4795	5	US-08-901-710-3
12	41.2	10.7	9108	6	PCT-US95-04228-45
13	41	10.6	1693	5	US-09-320-878-23
14	36.2	9.4	1931	4	US-09-130-114-2
15	35.6	9.2	1136	2	US-08-106-981-5
16	35.6	9.2	1431	1	US-08-254-357-1
17	35.6	9.2	1632	1	US-07-959-941-1
18	35.6	9.2	1632	1	US-08-259-924-1
19	35.6	9.2	4983	1	US-08-472-358-1
20	35.6	9.2	4983	6	PCT-US92-05786A-1
21	35.4	9.2	68750	5	US-09-335-409-1
22	34.8	9.0	2238	2	US-08-742-011-1
23	34.6	9.0	1157	1	US-08-095-726-1
24	34.6	9.0	1157	1	US-08-095-726-3
25	34.6	9.0	1157	1	US-08-096-043-1
26	34.6	9.0	1157	1	US-08-096-043-3

27	34.6	9.0	1157	1	US-08-093-577-1	Sequence 1, Appli
28	34.6	9.0	1157	1	US-08-093-577-3	Sequence 3, Appli
29	34.6	9.0	1157	1	US-08-096-623A-1	Sequence 1, Appli
30	34.6	9.0	1157	1	US-08-096-623A-3	Sequence 3, Appli
31	34.6	9.0	50341	2	US-08-247-901C-1	Sequence 1, Appli
32	34.6	9.0	50341	4	US-09-075-904-1	Sequence 1, Appli
33	34.4	8.9	1513	1	US-08-314-309A-2	Sequence 2, Appli
34	34.4	8.9	1546	1	US-08-314-309A-3	Sequence 3, Appli
35	34.4	8.9	1621	2	US-08-722-001-13	Sequence 13, Appli
36	34.4	8.9	1776	2	US-08-722-001-29	Sequence 2, Appli
37	34.4	8.9	2140	1	US-08-334-698-1	Sequence 29, Appli
38	34.4	8.9	2140	1	US-08-228-932-1	Sequence 1, Appli
39	34.4	8.9	2140	2	US-08-468-939-1	Sequence 1, Appli
40	34.4	8.9	2140	3	US-08-406-855A-1	Sequence 1, Appli
41	34.4	8.9	2140	4	US-08-722-190-1	Sequence 1, Appli
42	34.4	8.9	2140	5	US-08-244-354-1	Sequence 1, Appli
43	34.4	8.9	2140	5	US-09-206-899-1	Sequence 1, Appli
44	34.4	8.9	2140	6	PCT-US95-04203-1	Sequence 1, Appli
45	34.4	8.9	3172	1	US-08-314-309A-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-012-030-1  
; Sequence 1, Application US/09012030  
; Patent No. 5912169  
; GENERAL INFORMATION:  
; APPLICANT: SCHMIDT, Ralf-Michael, STITT, Marc, SONNEWALD,  
; APPLICANT: Uwe  
; TITLE OF INVENTION: Transketolase  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage.  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/012,030  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/590,454  
; FILING DATE: 22-JAN-1996  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2629 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Nicotiana  
; FEATURE:  
; NAME/KEY: CDA  
; LOCATION: 60..2289  
US-09-012-030-1

Query Match 44.3%; Score 170.4; DB 4; Length 2629;  
Best Local Similarity 67.4%; Pred. No. 3.6e-37;  
Matches 240; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 4 ccggcaccaagcctgacatcattggttggtggccaccgctcgagctgagatcgaggca 63









APPLICANT: Lee, James M.  
APPLICANT: Matthews, William  
APPLICANT: Tsai, Siao Ping  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04228  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/222616  
FILING DATE: 04-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Wendy M. Lee  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 821P3PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4425 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US95-04228-31

Query Match 10.7%; Score 41.2; DB 6; Length 4425;  
Best Local Similarity 46.3%; Pred. No. 0.024;  
Matches 136; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 51 gagatcgccggaatcgccgagcagctgaagaaggaggagagcgtccgcgtcgc 110  
DB 2086 GACCTCTGTGACGAGCGGACTCGCTGGAGATGCGAGTGTGTCGGCCGAGCGCAC 2145  
QY 111 tcgttcgtctcctgggaactctttgatgagcagtcgagtgatgacaaaggagagcgtcctc 170  
DB 2146 GCGCCAGCATCGTGTGTACAAAGACGAGAGGCTGCTGGAGGAAAGTCTGGAGTCGAC 2205  
QY 171 cctgcagactcagcagagagatcagcagtcagagccgggtccactctcggctggcagaag 230  
DB 2206 TTGGCGGACTCCAAACCAAGAGCTGAGCATCCAGCGGTGCGGAGGAGGATGCGGACGC 2265  
QY 231 tagctcgagcccaagcagcagcattggcatcgacaagttcgccgagtgctctgcc 290  
DB 2266 TATCTGTGACAGCTGTGCAACGCAAGGCTGCGTCAACTCTCCGCCAGCGTGGCCGCG 2325  
QY 291 gggacgatctacaaggagtagtcacgctggagagcatcattgcaactgcc 344  
DB 2326 GAAGGCTCCGAGGATAAGGCAACCATGGAGATCGTATCTTGTGCGTACCGGC 2379

RESULT 10  
US-08-340-011-3  
Sequence 3, Application US/08340011  
Patent No. 5776755  
GENERAL INFORMATION:  
APPLICANT: Alitalo, et al.

TITLE OF INVENTION: FLT4, A NOVEL RECEPTOR TYROSIN KINASE  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/340,011  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/959,951  
FILING DATE: 09-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Thomas C.  
REGISTRATION NUMBER: 36,989  
REFERENCE/DOCKET NUMBER: 32267  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4795 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20..4111  
US-08-340-011-3

Query Match 10.7%; Score 41.2; DB 2; Length 4795;  
Best Local Similarity 46.3%; Pred. No. 0.024;  
Matches 136; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 51 gagatcgccggaatcgccgagcagctgaagaaggaggagagcgtccgcgtcgc 110  
DB 2075 GACCTCTGTGAGCGGACTCGCTGGAGATGCGAGTGTGTCGGCCGAGCGCAC 2134  
QY 111 tcgttcgtctcctgggaactctttgatgagcagtcgagtgatgacaaaggagagcgtcctc 170  
DB 2135 GCGCCAGCATCGTGTGTACAAAGACGAGAGGCTGCTGGAGGAAAGTCTGGAGTCGAC 2194  
QY 171 cctgcagactcagcagagagatcagcagtcagagccgggtccactctcggctggcagaag 230  
DB 2195 TTGGCGGACTCCAAACCAAGCTGAGCATCCAGCGGTGCGGAGGAGGATGCGGACGC 2254  
QY 231 tagctcgagcccaagcagcagcattggcatcgacaagttcgccgagtgctctgcc 290  
DB 2255 TATCTGTGACAGCTGTGCAACGCAAGGCTGCGTCAACTCTCCGCCAGCGTGGCCGCG 2314  
QY 291 gggacgatctacaaggagtagtcacgctggagagcatcattgcaactgcc 344  
DB 2315 GAAGGCTCCGAGGATAAGGCGAGCATGGAGATCGTATCTTGTGCGTACCGGC 2368

RESULT 11  
US-08-901-710-3  
Sequence 3, Application US/08901710  
Patent No. 6107046  
GENERAL INFORMATION:  
APPLICANT: Alitalo, Kari

Query Match	10.7%;	Score 41.2;	DB 6;	Length 9108;
Best Local Similarity	46.3%;	Pred. No. 0.03;		
Matches 136;	Conservative	0;	Mismatches 158;	Indels 0; Gaps 0;
51	gagatcgcgggcaatg	cgccacagc	gtgaagaaaggaggggaaacacaggtccgcgtctc	110
3018	GACCTCTCGTGAAC	GTGAGCGAC	TGCGTGGAGATGCAGTGTCTGGTGGCGGAGCGCAC	3077
111	tgcttgcgtctctggg	aactctttgatgagcagtcggtgatgatacaaggagagcgtctc	170	
3078	GCGCCAGCATCGT	TGTGTACAAACAGCAGAGAGCTCTGGAGGAAAAGTCTGGAGTCGAC	3137	
171	cctgcgcagctcacgc	gagatcagcatcgaggccgggtccactctctcggtcgcagaag	230	
3138	TTGGCGGCTCCAAC	CAGAAGCTGAGCATCCAGCGCGTTCGCGGAGGAGGATGTCGGGAGCGC	3197	
231	tacatcgagcccaagc	aagccattgaacatcgacaatttcgagcgcagtgactcctgcc	290	

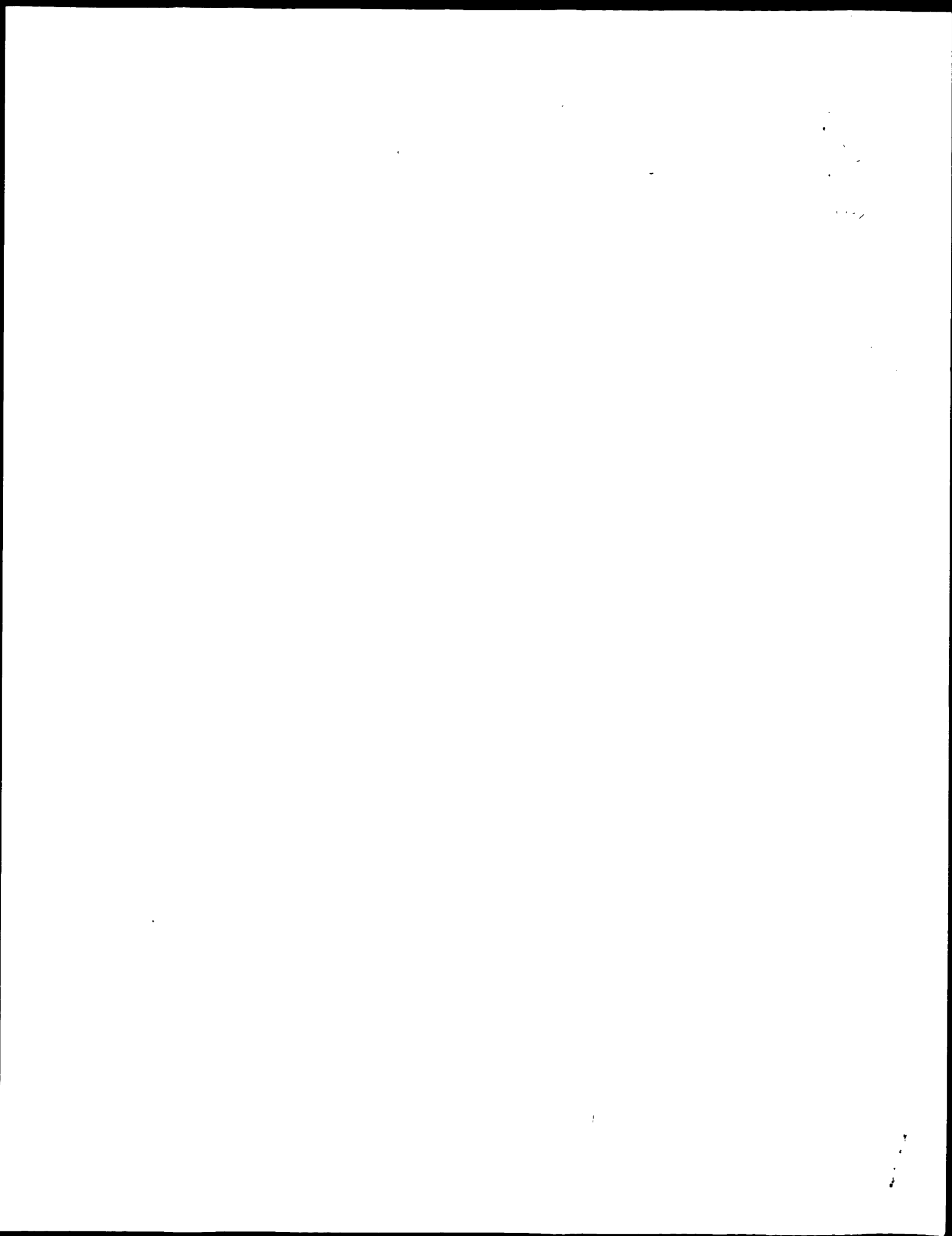




; STRANDEDNESS: Double  
; TOPOLOGY: Linear  
; MOLECULE TYPE: CDNA to mRNA  
; ORIGINAL SOURCE: Triticum aestivum L.  
US-08-106-981-5

Query Match 9.2% Score 35.6; DB 2; Length 1136;  
Best Local Similarity 51.2%; Pred. No. 0.49; 79; Indels 0; Gaps 0;  
Matches 83; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 4 ccggcaccagcctgacatcattgggttggtggcaccggtcccgagctggagatcgcgggca 63  
Db 130 CGCGCGAGCAGGTGGAGGTGACGGACAGGGGCTCTCTGGCAACCTCTCTGGCAAGAAGA 189  
QY 64 atgcgccgacgagctgaggaagaggggaagacgggtccgctgctgctgctgctgct 123  
Db 190 AGGCTGAGGAGGACAAAGGAGAGGAGGAGGAGCTGGTCACCGGATGGAGAGGTCTCCG 249  
QY 124 gggaaactctttagatgagcagtcgagatgagtagacagagagcg 165  
Db 250 TGGAGAGCCCCAGGTCAGGAAGGAGGAGGAGGACAGGAGGTGGCG 291

Search completed: November 4, 2000, 13:34:17  
Job time: 16693 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:33:25 ; Search time 4075.18 Seconds  
(without alignments)  
338.673 Million cell updates/sec

Title: US-09-300-482-225  
Perfect score: 316  
Sequence: 1 gataagtgccacactgag.....gaatggaggagactctga 316

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues 2067340  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_cm:\*

4: gb\_ov:\*

5: gb\_pat:\*

6: gb\_ph:\*

7: gb\_pl1:\*

8: gb\_pl2:\*

9: gb\_pr1:\*

10: gb\_pr2:\*

11: gb\_pr3:\*

12: em\_fun:\*

13: em\_hum1:\*

14: em\_hum2:\*

15: em\_in:\*

16: em\_om:\*

17: em\_or:\*

18: em\_ov:\*

19: em\_pat:\*

20: em\_ph:\*

21: em\_pl:\*

22: em\_ro:\*

23: em\_sts:\*

24: em\_sy:\*

25: em\_un:\*

26: em\_vi:\*

27: gb\_htg1:\*

28: gb\_htg2:\*

29: gb\_inl:\*

30: gb\_in2:\*

31: em\_ba1:\*

32: em\_ba2:\*

33: em\_hum3:\*

34: em\_hum4:\*

35: gb\_pr4:\*

36: gb\_htg3:\*

37: gb\_htg4:\*

38: gb\_htg5:\*

39: gb\_htg6:\*

40: gb\_htg7:\*

41: em\_htg1:\*

42: em\_htg2:\*

43: em\_htg3:\*

44: em\_hum5:\*

45: gb\_pl3:\*

46: gb\_pr5:\*

47: gb\_htg8:\*

48: gb\_htg9:\*

49: gb\_htg10:\*

50: gb\_htg11:\*

51: gb\_htg12:\*

52: gb\_htg13:\*

53: gb\_htg14:\*

54: gb\_in3:\*

55: gb\_htg15:\*

56: gb\_htg16:\*

57: gb\_htg17:\*

58: em\_htg4:\*

59: em\_htg5:\*

60: em\_htg6:\*

61: em\_htg7:\*

62: em\_hum6:\*

63: gb\_htg18:\*

64: gb\_htg19:\*

65: gb\_ba3:\*

66: em\_htg8:\*

67: em\_htg9:\*

68: em\_htg10:\*

69: em\_htg11:\*

70: em\_htg12:\*

71: em\_htg13:\*

72: em\_htg14:\*

73: em\_htg15:\*

74: em\_htg16:\*

75: em\_htg17:\*

76: em\_htg18:\*

77: em\_htg19:\*

78: em\_htg20:\*

79: em\_htg21:\*

80: em\_htg22:\*

81: em\_htg23:\*

82: gb\_pr6:\*

83: gb\_pr7:\*

84: gb\_htg20:\*

85: gb\_htg21:\*

86: gb\_htg22:\*

87: gb\_htg23:\*

88: gb\_ro:\*

89: gb\_sts1:\*

90: gb\_sts2:\*

91: gb\_sy:\*

92: gb\_un:\*

93: gb\_vil:\*

94: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	160.8	50.9	930	7	AF189365	AF189365 Oryza sat
2	78.4	24.8	89154	8	ATAC010797	AC010797 Arabidops
3	69.8	22.1	67712	8	ATAC022355	AC022355 Arabidops
4	68.4	21.6	5348	45	S77133	S77133 GRF1-genera
5	63	19.9	720	8	CNS01918	AL111059 Botrytis
6	60.6	19.2	2913	45	ZMU29162	U29162 Zea mays cl
7	57.8	18.3	424	89	G37533	G37533 SHGC-57943
8	54.8	17.3	1574	82	HSJA74326	AJ224326 Homo sapi
9	53.8	17.0	10304	7	AF166527	AF166527 Zea mays
10	52.2	16.5	3846	7	AB031012	AB031012 Zea mays
11	51.4	16.3	26292	40	AC019942	AC019942 Drosophila
12	51.4	16.3	27521	30	CELF08F8	U28991 Caenorhabdi

Sat Nov 4 18:11:12 2000

```

13 51.4 16.3 50089 29 AC005641
14 51.4 16.3 125150 27 AC005450
15 51.4 16.3 131899 27 AC008258
16 51.4 16.3 259764 29 AE003840
17 50.4 15.9 122925 85 AL360001
18 49.8 15.8 119907 27 AC008259
19 47.6 15.1 911 45 SCPOS18
20 47.6 15.1 1328 45 SCYJL121C
21 47.6 15.1 39729 45 SPAC31G5
22 47 14.9 5537 45 ZM2MM1
23 46.4 14.7 1127 83 HUMARS1H
24 45.6 14.4 182756 7 AC007789
25 44.6 14.1 82101 10 AC005694
26 44.6 14.1 126138 82 HSAY18000
27 44.6 14.1 149308 9 AC005527
28 44.6 14.1 153804 27 AC006280
29 44.6 14.1 196149 27 AC004709
30 44.6 14.1 318488 9 AC005529
31 44.2 14.0 1050 7 AF047444
32 44.2 14.0 135209 46 HS42616
33 44 13.9 7753 45 ZMAYSPG
34 43.4 13.7 8291 45 ZM057899
35 42.4 13.4 7753 45 ZMAYSPG
36 41.8 13.2 6414 45 ZMGPAL
37 41.8 13.2 6414 45 STEPELMR
38 41.2 13.0 1055 45
39 39.6 12.5 37262 29 AC009805
40 39.6 12.5 39475 29 AC016024
41 39 12.3 182756 7 AC007789
42 38 12.0 1307 2 MAGHPS
43 37.8 12.0 5537 45 ZM2MM1
44 37.6 11.9 139510 39 AC016779
45 37.2 11.8 10304 7 AF166527

ALIGNMENTS

RESULT 1
AF189365 930 bp mRNA 10-JAN-2000
LOCUS Oryza sativa D-ribulose-5-phosphate 3-epimerase mRNA, complete cds.
DEFINITION AF189365
ACCESSION AF189365
VERSION AF189365.1 GI:6007802
KEYWORDS Oryza sativa.
SOURCE Oryza sativa.
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 930)
Kopriya,S., Koprivova,A. and Suss,K.H.
Identification, cloning, and properties of cytosolic
D-ribulose-5-phosphate 3-epimerase from higher plants
J. Biol. Chem. 275 (2), 1294-1299 (2000)
JOURNAL 20092904
MEDLINE 2 (bases 1 to 930)
Kopriya,S., Koprivova,A. and Suss,K.H.
Direct Submission
Submitted (20-SEP-1999) Institute of Forest Biology and Tree
Physiology, Am Flughafen 17, Freiburg im Breisgau 79085, Germany
JOURNAL Location/Qualifiers
FEATURES
source
1..930
/organism="Oryza sativa"
/db_xref="taxon:4530"
35..721
/EC_number="5.1.3.1"
/note="cytosolic isoform"
/codon_start=1
/product="D-ribulose-5-phosphate 3-epimerase"
/protein_id="AA01048.1"
/db_xref="GI:6007803"
/translation="MAAARAAKIAPSMSSDFANLAAFAADRVRLGADWLHMDIMDGH
FVPLNTIGAPVIOSLRKRTKAYLDCHLMVTNPDSYVEFLAKAGASGTFTHIEVSRDNK

AC005641 Drosophil
AC005450 Drosophil
AC008258 Drosophil
AE003840 Drosophil
AL360001 Homo sapi
AC008259 Drosophil
X83571 S.cerevisia
Z49396 S.cerevisia
Z98979 S.pombe chr
X81199 Z.mays ZMM1
J04742 Human auton
AC007789 Oryza sat
AC005694 Homo sapi
Y18000 Homo sapien
AC005527 Homo sapi
AC006280 Plasmodiu
AC004709 Plasmodiu
AC005529 Homo sapi
AF047444 Oryza sat
AL020997 Human DNA
Z11879 Z.mays P ge
U57899 Zea mays re
Z11879 Z.mays P ge
M18976 Zea mays ch
X15408 Maize Gpal
Z50098 S.tuberosum
AC009805 Leishmani
AC016024 Leishmani
AC007789 Oryza sat
X81199 Z.mays ZMM1
AC016779 Oryza sat
AF166527 Zea mays

Query Match 50.98; Score 160.8; DB 7; Length 930;
Best Local Similarity 83.28; Pred. No. 1.6e-41;
Matches 183; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 gataaggtgcgcacactgagaaagaagaccctcccttcacatagaggttgatggtggt 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 518 GAAAGAGTGCCTGCCTCAGAAAAGAGTACCATCCCTTGACATCGAGGTTGATGGTGT 577
QY 61 ctagtctctcaacatcagatgagcgtggcgcacatctgctggggccaattgcctcgtgga 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 578 CTGGGTCTTCACCATGATGATGGCTGCATCTGCTGGGGCTAACTGCTGCTCTGGA 637
QY 121 agctctatttggcgtgcgagccagcagcagcattctgtgctgaggaagagcgtc 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 638 ACTCAATATTGGAGCAGCTGAACACAGGAGGTCATATCGGCACCTAAGGAAGCGCTT 697
QY 181 gagggctctcagaacaaactgatttgggtgttctgct 220
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 698 GAAGGATCTCAGAACAAAAGCTGATCTGTGTTTGTACT 737

RESULT 2
ATAC010797/c 89154 bp DNA PLN 09-DEC-1999
LOCUS Arabidopsis thaliana chromosome III BAC F28J7 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC010797
VERSION AC010797.2 GI:6091712
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 89154)
Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.,
Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E.,
Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome III BAC F28J7 genomic sequence
Arabidopsis thaliana chromosome III BAC F28J7 genomic sequence
Unpublished
JOURNAL 2 (bases 1 to 89154)
Lin,X. and Kaul,S.
Direct Submission
Submitted (23-SEP-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org
JOURNAL 3 (bases 1 to 89154)
Lin,X.
Direct Submission
Submitted (21-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
JOURNAL 4 (bases 1 to 89154)
Lin,X.
Direct Submission
Submitted (09-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
JOURNAL On Oct 21, 1999 this sequence version replaced gi:5919288.
COMMENT Address all correspondence to:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
E-mail: xlin@tigr.org
BAC clone F28J7 is from Arabidopsis chromosome III and is near the
molecular marker ILRI.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.

```

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://www.cds.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tadb/at/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

## FEATURES

## source

1. 89154

/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="III"  
/map="ILRI"  
/clone="F28J7"

## misc\_feature

1. 28249

/note="Overlap with BAC clone F4P13  
(AC009325:77296..105543)."

## misc\_feature

complement(784..815)

/note="exon predicted by xgrail, quality

marginal\_shadowexon"

join(<1449..1750,2015..2051,2153..2389,2484..2531,

2614..3264)

/gene="F28J7.1"

<1449..3264

/note="F28J7.1"

/gene="F4P13.23; similar to putative protein GB:CAA22987

[Arabidopsis thaliana]

join(1449..1750,2015..2051,2153..2389,2484..2531,

2614..3075)

/gene="F28J7.1"

/note="unknown protein"

/codon\_start=1

/protein\_id="AAF03425.1"

/db\_xref="GI:6091713"

/translation="MGVTSVAAKFAFPSPPSYKVVTDELTLGLLLSPFFHRENV  
EIVKLRGRGTEIVGMVVRHPMTATLLYSHGNAADIGOMYELFIELSIHLKYNLMGY  
DYSGYGQSTGPKSPENYADIEAVYKLEETFGSKQEGVILYGVSGSGPTLDLMSRL  
POLRAVLIHSILSLRVNYSVKVTKYWFIDYKNTDKIPYVDCPVLIIHGPTSDVDCS  
HGQLWELCKDKYEPPLWVGNGHCDLHYPEYIPHLKFIATVERLPCPRMSDQSER  
VRDAPPRRSMDRRYKPRQSTERREREKPKPSQKSSSSSKLISFDQLDRSSRSDC  
HEKTRKSDVQIERGKSVDRDLRVRSE"

<3677..4136

/gene="F28J7.2"

<3677..4136

/gene="F28J7.2"

/note="F4P13.24"

3677..4087

/gene="F28J7.2"

/note="unknown protein"

/codon\_start=1

/protein\_id="AAF03426.1"

/db\_xref="GI:6091714"

/translation="MARLFVVALLVAGTVFAADAPSAAPTASTPKSPTKAPAAAP  
KSSAAAPKASPAVEPTPEDDYSAPSPSACAPTVSSPAPTPEADGPFSSDGPSSD  
GPAAEPSKSGATNVKLSTAGIVAAAGFFIFSL"

4330..4401

/gene="F28J7.3"

/product="tRNA-Pro"

gene  
/anticodon="(pos:4362..4364,aa:Pro)  
4330..4401  
/gene="F28J7.3"  
/note="F4P13.25"  
complement(join(<4853..4932,5366..5690,5785..6288,  
6384..>6434))  
/gene="F28J7.4"  
complement(<4853..>6434)  
/gene="F28J7.4"  
/note="F4P13.26; predicted by genefinder"  
complement(join(4853..4932,5366..5690,5785..6288,  
6384..6434))  
/gene="F28J7.4"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="AAF03427.1"  
/db\_xref="GI:6091715"  
/translation="MGESACLMQPFYSYAAPOGDSLGALQSVSFGFRMSEKLDEKWS  
AFSTQPYVAEARYSKPGSAQKKAFFEAHYKVAARAKAAAEALLQOQIPKPOP  
EKSPQVEDINGEKKDMVSKPDLEIPRGLDVSQALKAELAEKGRFSNQSEKEN  
CGMAESEINGRTAEQVDEKPIILSKVKKSKDSQPKSSTPRVSKYNSERTPSKRS  
NNKSSSYTTPAKEFNRLYSIIIRKIDGSRASSKPKRDKTLPRTVSKNTAFKLCFT  
QPSAYALVMIAGVCKGNEARGEIQNGDTEPEGGECEGEGRE"  
complement(join(<7108..7719,7796..8075,8138..8302,  
8384..8859,8940..9140,9249..9492,9578..9748,  
10009..>10304))  
/gene="F28J7.5"  
complement(<7108..>10304)  
/gene="F28J7.5"  
/note="F4P13.27"  
complement(join(7108..7719,7796..8075,8138..8302,  
8384..8859,8940..9140,9249..9492,9578..9748,10009..>10304))  
/gene="F28J7.5"  
/note="unknown protein"  
/codon\_start=1  
/protein\_id="AAF03428.1"  
/db\_xref="GI:6091716"  
/translation="MRWDLITAIVAALVSVLADSGQMAYPIRHTLFSVEQNYFOW  
QTVGLMHSFLKSGQGPPIIRLLSCTDDQKTYIRGMNLAPTFEVSWSRHPKTDGWYPA  
INKPVGVLVWLQHSSEAKHVDVWVILDDAMIRGPIIPWELGAEGRPFPAHYGLVG  
CNLLVRLHTKHPCLDKVGLLAMDILRLVLAFLSKTEDVRODTAHTWNTLTDG  
IYKGWISEMYSGEAGLKKINDDLMIYGYVPRGVEPVLHMYGLPFSIGNM  
SFTKLDHEDNIVDCNRLFPEPPYREVKIMEPPDKRRGLLILSECMNTLNEGLIL  
RHAENGCPKAWTKYLSPLKSMETLTPKLLAPGSVHILPDQHEPPIDEFKGTY  
KIHLTFSECTYFDWQTVGFHSGFRQSGQPGNTIRLLSCTDEALNKGHDLPATY  
RGRVSTYSPSLKPSLFLSYLIGDNDLRLHTRNPEACDKVGGVLIIMHIEDLRK  
FAMYVLLKTOVRADKHYGKELTGDIVESGWISEMISYSGFAGAEELNLRSHKEIMI  
YGVYVPEGADYRVHGYGLEKFGVGNFWSFDRANRNTDLINKWAKFPDPPPSAVHOT  
DNDLRQDLLEIGQKLINEALFLHRRKNCPEPSETEKISVSRKVGNIETKQOG  
SDTKESGSGSESGRSTLKLWLVIALWLVISGVFLVMLLVSTRGRGTRTKGKYR  
NKRTSYSNFGFLDTK"  
complement(10410..10481)  
/note="exon predicted by xgrail, quality good"  
10571..10626  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
complement(10844..10930)  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
complement(11233..11268)  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
complement(11324..11357)  
/gene="F28J7.6"  
/note="F4P13.28; predicted by genscan"  
11324..11587  
/gene="F28J7.6"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="AAF03429.1"

## misc\_feature

complement(10410..10481)

/note="exon predicted by xgrail, quality good"

10571..10626

/note="exon predicted by xgrail, quality

marginal\_shadowexon"

complement(10844..10930)

/note="exon predicted by xgrail, quality

marginal\_shadowexon"

complement(11233..11268)

/note="exon predicted by xgrail, quality

marginal\_shadowexon"

<11324..>11587

/gene="F28J7.6"

<11324..>11587

/gene="F28J7.6"

/note="F4P13.28; predicted by genscan"

11324..11587

/gene="F28J7.6"

/note="hypothetical protein"

/codon\_start=1

/protein\_id="AAF03429.1"

Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown proteins'. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://www.repeatmasker.sourceforge.net/>). tRNA genes are identified by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://www.repeatmasker.sourceforge.net/>). tRNA genes are identified by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://www.repeatmasker.sourceforge.net/>).

```

Location/Qualifiers
1. .67712
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="mi1353"
/clone="F9N12"
1. .12863
/note="overlap with BAC clone F2K11
(AC008047:88005..100867)."
```

1

submitted sequence due to an overlap with another bac  
(F2K11)"

Join(3453. 4024,4432. 4526,4569. 4721,4795. 4877,  
4969. 5104,5177. 5350,5443. >5622)

/gene="F9N12.1"

3453 >5622

/gene="F9N12.1"

/note="contains similarity to Flavin-containing  
monooxygenase"

Join(3469. 4024,4432. 4526,4569. 4721,4795. 4877,  
4969. 5104,5177. 5350,5443. 5622)

/gene="F9N12.1"

/note="unknown protein"

/codon\_start=1

/protein\_id="AAF34845.1"

/db\_xref="GI:6997183"

/translation="MAPALSPTRSHHVVAVIGAGPAGLVAARELRREGHSVVVFEKQKQ  
VGSTWTYDEVSDPLSGVDRTRSVHSHVVRYSRLINGRECTGYRDFPFVVRSGVSR  
RRPFSHGVLAYLKDFAKEFGTEEMVRPTVEVKKVSPAAEEGFKWRIESTKEKKYK  
RDEIDYAVVCGNHVYERPLAQIIGLISWPGMGKSHSHYRIPEPRDKVLIFIDHNW  
AVLIGNSAAEDLSRIARVAKVHVACRSNPADTFIKOTGYNNLWTHSMITESVHEDQ  
SVYQWKTISVDIIMHCTGYKHFPLDITGVTVDDNRGPLYKDFVFPAPAFAPLS  
FTGIPQWLPFPMFELQSGKIAGVGLRIPLPKSDMMIEIKTFTSYTEVGQIGPKRYT  
HRMGTFQFYDNWLASQCGCSETEWRKEMCLANGVRKEAHPETRYDEWDDHHLVSE  
YDFSLS"

<6658. >9312

/gene="F9N12.2"

<6658. >9312

/gene="F9N12.2"

/note="contains similarity to NBS/LRR disease resistance  
protein"

6658. 9312

/gene="F9N12.2"

/note="unknown protein"

/codon\_start=1

/protein\_id="AAF34846.1"

/db\_xref="GI:6997184"

/translation="MGISFSPDPCKVNSQWLDMKVSYYTHNLKLNLALEKTKMEK  
KAKRDLRLRKREAEGLRDSFOFWLDSVAATVEDIITLLDRNNVEIORLCUCRNI  
CSKSLTSYRIGKSVFLRLREVEKLGKEVFVITEQASTAFEPERLPLOTIVGQDTMI  
DKAGHLMDPVGIMGVGMGVGKTTLLTQLYNNFNKDKCGFDIGIWNVYDLAEIGVPI  
KVQDEIAKQLGLEDQWTDKSKQGLCYNLIREKSEVLDDIWEKVDLAEIGVPI  
PRTKGRKLAFTTRSQVCARMGVHPHNEVQCLLENVAFDLQFKVGGTTLTSDGPI  
QLARIYAKKCGGLPALNLTIGETMSKRTIOEWRHAIHVLNSYAAEFTGEMDKVLPL  
KYSYDNLKGEQVSKSLYCALYPEDAKTLKEDLTHTWICEETIDGSEGETEKAEDKGY  
ITGKLVNRSLLMDDGGRVACHMDEVREMAJLWIASELGTQKEAFIVRACVGVRE  
PICKLVNVRNRMSSLEGNKIIHHLVSYCECMELITLLGREGYSIRSQLTISSEFREN  
MPKLAVDLSHNSLSELPSELSVLNLLYTEISHLPKGLQELKKIITHNLNE  
TRKLESTIGTSSHLNKLKILFRSLRPDLNLTVEKLEITHELEIITIDPRKALVE

/db_xref="GI:6091717"	
/translation="MRASFLMICVLVLTALIMVAESKSSVSGNDETLPPTPPMPATSPSP	
PTGTGGHGFGPEKLLPPNNAFTFTCLITSLISALAFV"	
12299..12408	
/note="exon predicted by xgrail, quality marginal"	
<12814..13199	
/gene="F28J7.7"	
<12814..13199	
/gene="F28J7.7"	
/note="F4P13.29"	
12814..13194	
Query Match	
Best Local Similarity 24.8%; Score 78.4; DB 8; Length 89154;	
Matches 109; Conservative 0; Mismatches 51; Indels 0; Gaps 0;	
QY	47 aggttgatggtgctaggtctcctaaccatagagtgccgacatctctctgggccaatt 106
DB	45282 AGGTGGATGGCGGCTTAGGCCCTTCAACGATGCACACAGCGCTGCACGACGGCCCAACT 45223
QY	107 gcatcgctgcgtgaagctctctatttggcgctgcgagccagagccatcatctctgtgc 166
DB	45222 GTATTGTTGCTGGAGTTCAGTCTTTGGAGCTCGGAGCCAGGAGATGTCATCTCCCTTC 45163
QY	167 tgaggaagagctgcgagggctctcagaacaaactgatt 206
DB	45162 TGAGGACCAGTGTGTGAGAAAGCCACCACCTTGAGT 45123
RESULT 3	
ATAC022355	67712 bp DNA PLN 18-FEB-2000
LOCUS	Arabidopsis thaliana chromosome I BAC F9N12 genomic sequence,
DEFINITION	complete sequence.
ACCESSION	AC022355
VERSION	AC022355.3 GI:6997182
KEYWORDS	HTG.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Arabidopsis thaliana
AUTHORS	Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B., Wu,D., Matz,R., Rensing,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Niernan,W.C. and Fraser,C.M. I; Brassicaceae; Arabidopsids.
TITLE	1 (bases 1 to 67712)
JOURNAL	2 (bases 1 to 67712)
REFERENCE	Unpublished
AUTHORS	Lin,X. and Kaul,S.
TITLE	Direct Submission
JOURNAL	Submitted (01-FEB-2000) The Institute for Genomic Research, 9712
REFERENCE	Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
AUTHORS	3 (bases 1 to 67712)
TITLE	Lin,X.
JOURNAL	Direct Submission
REFERENCE	Submitted (18-FEB-2000) The Institute for Genomic Research, 9712
AUTHORS	Medical Center Dr., Rockville, MD 20850, USA
COMMENT	On Feb 18, 2000 this sequence version replaced gi:5910594. Address all correspondence to:at@tigr.org
BAC clone F9N12 is from Arabidopsis chromosome I and is near the	
molecular marker mi353.	
The orientation of the sequence is from SP6 to T7 end of the BAC	
clone.	
Genes were identified by a combination of three methods: Gene	
prediction programs including GRAIL (available by anonymous ftp	
from arturh.epm.ornl.gov), Genefinder (Phil Green, University of	
Washington, GenScan (Chris Burge,	
http://gnomic.stanford.edu/~chris/genSCANW.html), and NetPlantGene	
(http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the	
complete sequence against a peptide database and the Arabidopsis	
EST database at TIGR (http://www.tigr.org/tldb/at/at.html).	



	BASE COUNT	1325 a	1169 c	1203 g	1651 t	
	ORIGIN					
	Query Match		21.6%; Score 68.4;	DB 45; Length 5348;		
	Best Local Similarity		87.2%; Pred. No. 3.2e-11;			
	Matches	75; Conservative	0; Mismatches 11;	Indels 0;	Gaps 0;	
Qy	225	agtaactccctcgcttttttttatttcgcgtgcttttagttcaaacatgaacatacgcggagc	284			
Db	3413	AATACCTCCCTCGTGTGTTTATTGTCACAGTTTAGTATAAATAAACTAGCGGGC	3472			
Qy	285	actgatctcagaagtggaggagta	310			
Db	3473	ACAAATATTCGAGAATGGAGGTAGTA	3498			
	RESULT	5				
	CNS01918					
	LOCUS	720 bp	mRNA	PLN	02-SEP-1999	
	DEFINITION	Botrytis cinerea strain T4 cdna library under conditions of nitrogen deprivation.				
	ACCESSION	AL111059				
	VERSION	AL111059.1	GI:5825679			
	KEYWORDS	cdna library; nitrogen deprivation.				
	SOURCE	Botryotinia fuckeliana				
	ORGANISM	Eukaryota; Fungi; Ascomycota; Euascomycetes; Discomycetes; Botryotinia fuckeliana				
	REFERENCE	1 (bases 1 to 720)				
	AUTHORS	Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.				
	TITLE	Direct Submission				
	JOURNAL	Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France				
	REFERENCE	2 (bases 1 to 720)				
	AUTHORS	Genoscope.				
	TITLE	Direct Submission				
	JOURNAL	Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
	COMMENT	- Web : www.genoscope.cns.fr) The cdna library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.				
	FEATURES	Location/Qualifiers				
	source	1..720				
	/organism="Botryotinia fuckeliana"					
	/strain="T4"					
	/db_xref="taxon:40559"					
	/note="Genoscope sequence ID : W62E091"					
	BASE COUNT	226 a	146 c	178 g	170 t	
	ORIGIN					
	Query Match		19.9%; Score 63;	DB 8; Length 720;		
	Best Local Similarity		65.08; Pred. No. 1.3e-09;			
	Matches	93; Conservative	0; Mismatches 50;	Indels 0;	Gaps 0;	
QY	3	taagggtgcgcacactgagaagaagacccttccttgacatagaggttgatgggtgtct	62			
Db	578	TAAGGTCCGCAGACTCAGAAACGATACCACCACTCAACATCATGAAGTTGATGGTGATT	637			
QY	63	aggtctctcaacctagacgtggccgcatctgctggggccaattgccatctcgctggaag	122			
Db	638	GGGTCTGGAACAATGTATCAAGCCGAGATGCTGGCGCAATGTGTTGGTCAAAGCAG	697			
QY	123	ctctatatgttgccgtggcgacc	145			
Db	698	TGCTGTTTTTGGAGCCAAGGACC	720			





[illegible]

```

Db 7668 AAGTGTACTTCCATCTCTTTTATTGTCGTTTTTATTCAAAATGAACATAGTAG 7727
QY 283 CQACTGATTCGAGAGTGGAGGAGTACTT 313
Db 7728 CGACAAATATTCGAGAACTGAGGTAGTGT 7758

RESULT 10
AB031012/c
LOCUS AB031012 3846 bp DNA PLN 12-JUL-2000
DEFINITION Zea mays ZmRR2 gene for response regulator 2, complete cds.
ACCESSION AB031012
VERSION AB031012.1 GI:6009896
KEYWORDS response regulator 2.
SOURCE Zea mays (sub_species:Golden Cross Bantam T51) DNA.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
Deji,A., Sakakibara,H., Ishida,Y., Yamada,S., Komari,T., Kubo,T.
and Sugiyama,T.
Genomic organization and transcriptional regulation of maize ZmRR1
and ZmRR2 encoding cytokinin-inducible response regulators
Biochim. Biophys. Acta 1492, 216-220 (2000)
REFERENCE 2 (bases 1 to 3846)
AUTHORS Sakakibara,H., Deji,A. and Sugiyama,T.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) to the DDBJ/EMBL/GenBank databases. Hitoshi
Sakakibara, Nagoya University, Graduate School of Bioagricultural
Sciences; Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan
(E-mail:sakakibagr.nagoya-u.ac.jp, Tel:81-52-789-4105,
Fax:81-52-789-4107)
FEATURES
source
1. .3846
/organism="Zea mays"
/sub_species="Golden Cross Bantam T51"
/db_xref="taxon:4577"
2842. .2849
join(3000..3186,3113..3599)
/gene="ZmRR2"
join(3000..3186,3113..3599)
/gene="ZmRR2"
/note="cytokinin-inducible"
/codon_start=1
/product="response regulator 2"
/protein_id="BA085113.1"
/db_xref="GI:6009897"
/translation="MAAAATATPSVAPESGDRKAVPPVDVDAVLELEKHLAVD
SSVDRAVIKLRSSRYRTTVDSATRALELLALGLVPDVMIIITDYMGMGTGYELL
KHVKESSALRAIPVIMSSENVTRISRCLEGAEDFLKPRPADVSLCSRIIR"
3187. .3312
intron 975 a 883 c 971 g 1017 t
BASE COUNT 975 a 883 c 971 g 1017 t
ORIGIN

Query Match 16.5%; Score 52.2; DB 7; Length 3846;
Best Local Similarity 82.2%; Pred. No. 5.9e-06;
Matches 60; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 238 ttttttttttgcgcgttttagttcaaacatgaactgaactgcgagactgattcag 297
Db 1061 TCTTTTATTGTCGTCGTTAGTTTAAATGAATAGTGTAGTGAACGACAAATATTCGAG 1002
QY 298 aatgaggaggagta 310
Db 1001 AACGGAGATAGTA 989

RESULT 11
AC019942
LOCUS AC019942 26292 bp DNA HTG 03-JAN-2000
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered

```

```

pieces.
AC019942
VERSION AC019942.1 GI:6664955
HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 26292)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDW:10211468 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. .26292
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 7122 a 6111 c 5943 g 7116 t
ORIGIN

Query Match 16.3%; Score 51.4; DB 40; Length 26292;
Best Local Similarity 53.2%; Pred. No. 1.5e-05;
Matches 109; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 4 agagtgccacactgagaagaagtaacctccctccctgcacatgagagtggtggtcta 63
Db 24474 AAGGTCAAGTGGCTGCGCGAAAACTACCCCAACTGGACATCGAGTGGACGAGGTG 24533
QY 64 ggtcctcaaccatagacgtgcccgcactctgtggggccaattgcctgcgtggaagc 123
Db 24534 GGACCAAGACTATACACTGCTGTCGCCGAGCGCGACCAACATGATCTCTCGGAACC 24593
QY 124 tctatatgttcgctgcgagcaccaggagccatcatctctgtctgaggaagagcctcag 183
Db 24594 CGCGTGGTGGCGCTCCCGATCAGTCGCAGGTCATCAAGGAGTGGCGGATGTGTCG 24653
QY 184 ggcctcagaacaaaactgatttt 208
Db 24654 AGCTACCTCAATAGATTCATTT 24678

RESULT 12
CELFF08F8/c
LOCUS CELFF08F8 27521 bp DNA INV 15-JUN-1995
DEFINITION Caenorhabditis elegans cosmid F08F8.
ACCESSION U28991
VERSION U28991.1 GI:861364
KEYWORDS
SOURCE Caenorhabditis elegans strain-Bristol N2.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 27521)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Draxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Smith,A.,
Saunders,D., Showkeen,R., Smaiden,N., Smith,A., Sonhammer,E.,
Staden,K., Waterston,J., Thiermy-Mieg,J., Thomas,K., Vaudin,M.,
Wilkinson-Sproat,J. and Wohldman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```





Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

# REFERENCE AUTHORS

1 (bases 1 to 125150)  
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

## Sequencing of Drosophila melanogaster

### Unpublished

### REFERENCE

### AUTHORS

2 (bases 1 to 125150)  
Celniker, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,  
Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,  
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,  
Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,  
Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,  
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L.,  
Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,  
Zieran, L.L. and Kimmel, B.E.

### Direct Submission

### JOURNAL

Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley

### COMMENT

Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Jul 30, 1999 this sequence version replaced gi:5597051.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bagp@fruitfly.berkeley.edu](mailto:bagp@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases. Pl library location:  
7-33.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 49 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 610: contig of 610 bp in length

\* 611 690: gap of unknown length

\* 691 1308: contig of 618 bp in length

\* 1309 1388: gap of unknown length

\* 1389 1942: contig of 554 bp in length

\* 1943 2022: gap of unknown length

\* 2023 2583: contig of 561 bp in length

\* 2584 2663: gap of unknown length

\* 2664 3865: contig of 1202 bp in length

\* 3866 3945: gap of unknown length

\* 3946 4978: contig of 1032 bp in length

\* 4978 5057: gap of unknown length

\* 5058 5799: contig of 662 bp in length

\* 5799 5800: gap of unknown length

\* 5800 6452: contig of 652 bp in length

\* 6452 7403: gap of unknown length

\* 7403 7484: contig of 872 bp in length

\* 7484 8621: gap of unknown length

\* 8621 8702: contig of 1138 bp in length

\* 8702 9614: contig of 913 bp in length

\* 9615 9694: gap of unknown length

\* 9695 10745: contig of 1050 bp in length

\* 10745 10825: gap of unknown length

\* 10825 11537: contig of 633 bp in length

\* 11538 12289: contig of 752 bp in length

\* 12290 12370: gap of unknown length

\* 12370 14469: contig of 2100 bp in length

\* 14470 14532: gap of unknown length

\* 14533: contig of 776 bp in length

\* 14550

15326 15405: gap of unknown length  
15406 16338: contig of 933 bp in length  
16339 16418: gap of unknown length  
16419 17556: contig of 1138 bp in length  
17557 17636: gap of unknown length  
17637 20200: contig of 2564 bp in length  
20201 20280: gap of unknown length  
20281 21724: contig of 1444 bp in length  
21725 21804: gap of unknown length  
21805 22429: contig of 625 bp in length  
22430 22509: gap of unknown length  
22510 23332: contig of 823 bp in length  
23333 23412: gap of unknown length  
23413 24354: contig of 942 bp in length  
24355 24434: gap of unknown length  
24435 26644: contig of 2210 bp in length  
26645 26724: gap of unknown length  
26725 28963: contig of 2239 bp in length  
28964 29043: gap of unknown length  
29044 37172: contig of 8129 bp in length  
37173 37252: gap of unknown length  
37253 110032: contig of 72780 bp in length  
110033 110112: gap of unknown length  
110113 110721: contig of 609 bp in length  
110722 110801: gap of unknown length  
110802 111445: contig of 644 bp in length  
111446 111525: gap of unknown length  
111526 112124: contig of 599 bp in length  
112125 112205: gap of unknown length  
112206 112817: contig of 613 bp in length  
112818 112897: gap of unknown length  
112898 113527: contig of 630 bp in length  
113528 113607: gap of unknown length  
113608 114212: contig of 605 bp in length  
114213 114292: gap of unknown length  
114293 114906: contig of 614 bp in length  
114907 114986: gap of unknown length  
114987 115598: contig of 612 bp in length  
115599 115678: gap of unknown length  
115679 116267: contig of 589 bp in length  
116268 116347: gap of unknown length  
116348 116981: contig of 634 bp in length  
116982 117061: gap of unknown length  
117062 117399: contig of 338 bp in length  
117400 117479: gap of unknown length  
117480 118110: contig of 631 bp in length  
118111 118190: gap of unknown length  
118191 118836: contig of 646 bp in length  
118837 118916: gap of unknown length  
118917 119535: contig of 619 bp in length  
119536 119615: gap of unknown length  
119616 120248: contig of 633 bp in length  
120249 120328: gap of unknown length  
120329 120951: contig of 623 bp in length  
120952 121031: gap of unknown length  
121032 121644: contig of 613 bp in length  
121645 121724: gap of unknown length  
121725 122340: contig of 616 bp in length  
122341 122420: gap of unknown length  
122421 123036: contig of 616 bp in length  
123037 123116: gap of unknown length  
123117 123756: contig of 640 bp in length  
123757 123836: gap of unknown length  
123837 124448: contig of 612 bp in length  
124449 124528: gap of unknown length  
124529 125150: contig of 622 bp in length.

## FEATURES

### source

1. 125150  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/clone="DS00609 (D364)"  
/chromosome="2"







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:44 ; Search time 320.8 Seconds  
(without alignments)  
351.306 Million cell updates/sec

Title: US-09-300-482-619

Perfect score: 300

Sequence: 1 gattattgacacaccgat.....ttcagaaatgtctgtgtgg 300

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: N\_Geneseq\_36:\*

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.2	12.1	10820	20	X20527
C 2	32.4	10.8	543	18	V75242
C 3	31.6	10.5	32768	20	X13336
C 4	30.8	10.3	579	20	X87836
C 5	30.8	10.3	1004	20	X13479
C 6	30.6	10.2	1051	19	V27386
C 7	30.6	10.2	13206	19	V52166
C 8	30.6	10.2	22243	18	V74475
C 9	29.8	9.9	1062	20	Z39041
C 10	29.8	9.9	1190	19	V61936
C 11	29.8	9.9	1308	20	X25510
C 12	29.8	9.9	1373	19	V52969

13	29.8	9.9	1683	21	Z99543	Human insulin-resp
14	29.8	9.9	1750	20	V74136	Human FLAME-1 cDNA
15	29.8	9.9	1996	20	X25508	Human MACH related
16	29.8	9.9	2034	19	V44806	I-FLICE-1 coding s
17	29.8	9.9	2040	21	Z48769	Human CLARP coding
18	29.8	9.9	2045	19	V50436	Human Casper gene.
19	29.8	9.9	2064	21	Z99539	Human insulin-resp
20	29.8	9.9	2084	21	Z56987	Usurpin-alpha poly
21	29.8	9.9	2138	19	V24137	Homo sapiens B258
22	29.8	9.9	2143	19	V61937	Human CFLIP-L DNA.
23	29.8	9.9	2143	20	Z39040	Human FLICE-like i
24	29.8	9.9	2188	19	V31375	Human FIN-1 encodi
25	29.8	9.9	2243	19	V52968	Human G1 protein i
26	29.8	9.9	2352	21	Z47926	Human apoptosis as
27	29.8	9.9	12923	10	N90338	Sequence of human
28	29.6	9.9	4042	20	X20659	Polynucleotide seq
29	29	9.7	2297	19	V52836	Human iduronate-2-
30	29	9.7	2297	19	V52836	Human IDS cDNA. H
31	29	9.7	2297	20	X86181	cDNA encoding idur
32	29	9.7	4428	19	V52837	Human iduronate-2-
33	29	9.7	4428	19	V15700	Human IDS gene. H
34	29	9.7	4428	20	X86183	DNA encoding iduro
35	28.8	9.6	1716	20	Z22975	Brassica napus D22
36	28.8	9.6	13425	19	V52284	Streptococcus pneu
37	28.6	9.5	13188	19	V52203	Streptococcus pneu
38	28.6	9.5	1664976	19	V21209	Methanococcus jann
39	28.4	9.5	3872	19	V49601	Candida albicans O
40	28.2	9.4	10726	19	V52199	Streptococcus pneu
41	28	9.3	1932	21	A26946	Essential Staphylo
42	28	9.3	15218	19	V18275	RSV isolate 2B wil
43	28	9.3	15218	20	Z22909	Nucleotide sequenc
44	28	9.3	15218	20	Z22914	Nucleotide sequenc
45	28	9.3	15218	20	X35267	DNA encoding the L

#### ALIGNMENTS

RESULT 1  
X20527/c  
ID X20527 standard; DNA; 10820 BP.  
XX  
XX X20527;  
XX  
XX 05-MAY-1999 (first entry)  
XX  
XX Polynucleotide sequence from the genome of Treponema pallidum.  
XX  
XX Treponema pallidum infection; syphilis; Borrelia infection; animal;  
XX enzyme production; ds.  
XX  
XX Treponema pallidum.  
XX  
XX WO9859034-A2.  
XX  
XX 30-DEC-1998.  
XX  
XX 23-JUN-1998; 98WO-US13041.  
XX  
XX 24-JUN-1997; 97US-0050667.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Fraser CM;  
XX  
XX WPI; 1999-081273/07.  
XX  
XX New isolated Treponema pallidum nucleic acids - used to develop  
XX products for the detection, diagnosis, characterisation, prevention  
XX and therapy of T. pallidum infections, particularly syphilis  
XX  
XX Claim 1; Page 343-349; 1150pp; English.  
XX





(HUMA-) HUMAN GENOME SCI INC.

Barash SC, Dillon PJ, Kunsch CA;

WPI; 1999-045171/04.

New isolated Enterococcus faecalis polynucleotides and polypeptides  
used to develop products for the detection of Enterococcus and for  
use in vaccines for prevention or attenuation of Enterococcus  
infection.

Claim 1; Page 1797-1798; 2084pp; English.

A computer readable medium has been developed which has recorded on it  
982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
X12938 to X13919 represent these nucleotide sequences which are primary  
nucleotide sequences, also known as contigs. The computer-based system  
can identify fragments of the Enterococcus faecalis genome with  
commercial importance. The products can be used to detect the presence  
of Enterococcus faecalis in samples. They can also be used for  
diagnosing Enterococcal infection in an animal and monitoring  
progression of disease, and for identifying agents which can be used to  
modulate the growth or pathogenicity of Enterococcus faecalis, or  
another related organism, in vivo or in vitro. In particular the  
polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
can be used in vaccines to prevent or attenuate an Enterococcal  
infection.

Sequence 1004 BP; 304 A; 188 C; 187 G; 316 T; 9 other;

Query Match                  10.3%; Score 30.8; DB 20; Length 1004;  
Best Local Similarity      47.0%; Pred. No. 1.5;  
Matches    86; Conservative         2; Mismatches    95; Indels    0; Gaps    0;

QY    103 gcaacctgaaccgcgaatggtcactagaagtacagaaagcccttcagagatcggggc 162  
| | | | | | | | | | | | | | | | | | | |  
DB    788 gctcgaagcacccaccgcacactgcagtgatgcgaacgccatttttngtc 847  
  
QY    163 tgcaccttcgaaaagggttgtcaattactcaagaaaaattctctgttgataaacctg 222  
| : | | | | | | | | | | | | | | | | |  
DB    848 yttaactcagccactgtagttaaatyctcataaaattctcatcatcgtaaataat 907  
  
QY    223 ctagaatagaggatggttagctcggttcttatgttattggtggtgtaggactt 282  
| | | | | | | | | | | | | | | | | |  
DB    908 cttcataaagcatgatgatgtctgttcttctatcgtataaattctggaagtacacgt 967  
  
QY    283 cag 285  
| |  
DB    968 ctg 970

RESULT    6  
V27386/C  
ID   V27386 standard; DNA; 1051 BP.  
XX AC   V27386;  
XX DT  
XX 02-OCT-1998 (first entry)  
STreptococcus pneumoniae SP0075 nucleotide.  
Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
detection; pneumonia; otitis media; meningitis; ss.  
Streptococcus pneumoniae.  
XX Key   Location/Qualifiers  
PH CDS    2..1051  
FT       /tag= a  
FT       /product= "SP0075"  
FT note= "no stop codon given"  
XX "

PX	W09818930-A2.
XX	
PD	07-MAY-1998.
XX	
PF	30-OCT-1997; 97WO-US19422.
XX	
XX	31-OCT-1996; 96US-0029960.
PR	(HUMA-) HUMAN GENOME SCI INC.
XX	
PA	Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
PI	WPI: 1998-272224/24.
DR	P-PADB; W55125.
DR	
XX	Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT	pneumoniae - or their epitope-containing fragments, useful in
PT	protective or therapeutic vaccines, and for diagnosis
XX	
PS	Claim 1; Page 75-76; 118pp; English.
XX	
CC	The present sequence encodes a protein from Streptococcus pneumoniae.
CC	The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC	can be useful in vaccines for inducing protective antibodies against
CC	Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC	pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC	are used to detect Streptococcus infection (by usual hybridisation or
CC	amplification methods), also for isolating Streptococcus genes or their
CC	allelic variants. The protein can be used similarly to detect specific
CC	antibodies in standard immunoassays, especially for diagnosing or
CC	monitoring infections. Antibodies which bind the protein are used to
CC	detect corresponding antigens, to purify the protein and for passive
CC	immunisation (optionally coupled to a toxin). Vaccines are administered,
CC	e.g. by injection, orally or through the skin, typically at 0.01-1000
CC	(especially 10-300) mu g/ml per dose.
XX	
XX	Sequence 1051 BP; 283 A; 225 C; 263 G; 280 T; 0 other;
SQ	
Query Match 10.2%; Score 30.6; DB 19; Length 1051;	
Best Local Similarity 53.8%; Pred. No. 1-8;	
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps	
QY	12 aacaccgatcctgtgggattgatcatcaaatgctcaactaggaacctgaactggcaact 71 
DB	677 RATCCCACTCCCCTTCGGCATGACAAACAGGGTAATCTTCCATGTCTTTGGCAAGCTGAT 618
QY	72 actcttgtaattgcatttcttaagacggaggcacacctgaaccgccgaatggctcta 128 
DB	617 AATCTCGTAGTGTCCTTTTTTGAGACGGACATCCACTTATCATTAAAGGCTCTA 561
RESULT 7	
V52166	
ID	V52166 standard; DNA; 13206 BP.
XX	
AC	V52166;
XX	
DT	23-OCT-1998 (first entry)
XX	
DE	Streptococcus pneumoniae genome fragment SEQ ID NO:33.
XX	
KW	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW	computer readable medium; vaccine; pharmaceutical composition; ds.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	W09818931-A2.
XX	
PD	07-MAY-1998.
XX	
PF	30-OCT-1997; 97WO-US19580.
XX	

```
PR 31-OCT-1996; 96US-0029960.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
XX Kunsch CA, Rosen CA;
XX WPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
XX
XX Claim 1; Page 332-339; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S.pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX to 391, identifying members of the library which contain sequences
XX that hybridise to the target sequence and isolating the nucleic acid
XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX from an organism, amplifying nucleic acid molecules whose nucleotide
XX sequence is homologous to amplification primers derived from the
XX fragment of the S. pneumoniae genome to prime the amplification and
XX isolating the amplified sequences. The computer readable medium can be
XX used in a computer-based system for identifying fragments of the
XX S. pneumoniae genome of commercial importance, or expression modulating
XX fragments of the S. pneumoniae genome. Products from the present
XX invention can be used in diagnosis kits and assays, and pharmaceutical
XX compositions and vaccines for S. pneumoniae.
XX
XX Sequence 13206 BP; 4016 A; 2929 C; 2473 G; 3786 T; 2 other;

Query Match 10.28; Score 30.6; DB 19; Length 13206;
Best Local Similarity 53.8%; Pred. No. 4.8;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 12 aacaccgatcgtcgtggattgatcatcaaatgctcaactagtagacgtggaact 71
|| || || || || || || || || || || || || || || || || || ||
Db 9898 aatccactccctctcgtgcatgacacagggttaattcttcattgcttggaagctgt 9957

Qy 72 actcttgtaattgtcattcttaagcgcggagggcacacctgaaccccgcaatgctcta 128
| ||| ||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9958 aatctcgtagtgtccctttttgagcagcgacatccattcataacctaaggctcta 10014

RESULT 8
W74475
ID V74475 standard; DNA; 22243 BP.
XX
XX V74475;
XX
XX 16-MAR-1999 (first entry)
XX
XX Staphylococcus aureus contig SEQ ID #164.
XX
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
XX
XX Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH
```











CC overlapping EST clones were sequenced to deduce the sequence of  
CC the full-length MRIT alpha 1 clone and 2 isoforms, alpha 2 (see  
CC X25509) and beta 1 (see X25510). The MRIT gene was localised to  
CC the human chromosome 2q32-33 region. MRIT alpha 1 includes an  
CC N-terminal death effector domain and a C-terminal caspase homology  
CC domain, but is not a cysteine protease. Selective enhancers and  
CC inhibitors of MRIT apoptotic activity can be identified and used  
CC to treat diseases mediated by the dysfunction of programmed cell  
CC death or proliferation. In addition, nucleic acid molecules  
CC encoding pro-apoptotic isoforms of MRIT can also be used as  
CC inducers of programmed cell death. A cell accumulation disorder  
CC such as cancer, autoimmune disease, viral infection, angiogenesis  
CC and atherosclerosis is treated by administering an agent that  
CC selectively enhances MRIT apoptotic activity, thereby inducing  
CC apoptosis in a subject. A disorder of cell loss, such as a  
CC neurodegenerative disorder, including Alzheimer's disease,  
CC Parkinson's disease, retinitis pigmentosa, stroke, aplastic  
CC anaemia, myocardial infarction or AIDS can be treated by  
CC administering an agent that selectively inhibits MRIT apoptotic  
CC activity.

XX Sequence 1996 BP; 543 A; 461 C; 540 G; 451 T; 1 other;

Query Match 9.9%; Score 29.8; DB 20; Length 1996;  
Best Local Similarity 51.1%; Pred. No. 4.2;  
Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
Qy 54 ggacctgaactggcaactactctgttaattgtcattcttaagagcgaggcacacctgaa 113  
Db ||| ||| ||| | || || | ||||| | |||||  
387 ggagctgtactgcaagacccttgagcttccctagcttaagataggtgtctgtgaa 446  
Qy 114 accggaatggtctactagaagtacagaagccttcagagatcgggggtgcaattctcg 173  
Db | | | | | ||||| || | | | ||||| |||  
447 gtatccatcagggttgaaagcacttgatcacagatgagaaggagatgctgctctttg 506  
Qy 174 aaacagggtgttgcaat 190  
Db | | | ||||| ||  
507 tgccgggatgtgtctat 523

Search completed: November 4, 2000, 13:45:55  
Job time: 16482 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:34:22 ; Search time 189.35 Seconds  
(without alignments)  
239.622 Million cell updates/sec

Title: US-09-300-482-619  
Perfect score: 300

Sequence: 1 gattattgacaacacgat.....ttcagaaatgtctgtgtgg 300

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*

7: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29.8	9.9	1750	5	US-08-859-167-1
2	29.8	9.9	1750	5	US-09-109-273-1
3	29	9.7	2297	2	US-08-484-493-1
4	29	9.7	2297	2	US-08-484-493-1
5	29	9.7	2297	2	US-08-484-493-1
6	29	9.7	2297	2	US-08-484-493-1
7	29	9.7	4428	2	US-08-484-493-6
8	29	9.7	4428	2	US-08-484-493-6
9	28	9.3	1932	5	US-08-345-212-6
10	28	9.3	1932	5	US-08-345-212-6
11	27.6	9.2	1146	3	US-08-714-918-97
12	27.6	9.2	1146	3	US-08-892-403A-2
13	27.6	9.2	3906	2	US-08-487-810-1
14	27.6	9.2	4165	2	US-08-469-537A-102
15	27.4	9.1	1055	2	US-08-442-248-1
16	27.4	9.1	1055	2	US-08-440-815-1
17	26.8	8.9	1055	6	US-08-618-464-1
18	26.8	8.9	2441	1	US-08-332-312-1
19	26.8	8.9	2624	1	US-08-032-382B-1
20	26.8	8.9	3705	6	PCT-US96-03940-7
21	26.4	8.8	5648	6	PCT-US96-03940-1
22	26.4	8.8	1806	4	US-08-980-060-12
23	26.4	8.8	3089	1	US-08-980-060-1
24	26.4	8.8	3459	4	US-07-921-796-7
25	26.4	8.8	6030	2	US-08-980-060-3
26	26.4	8.8	8700	3	US-08-441-139-8
					US-08-392-625-16

27	26.4	8.8	8700	3	US-08-466-961A-16	Sequence 16, Appl
28	26.4	8.8	8700	4	US-08-645-193B-18	Sequence 18, Appl
29	26.4	8.8	11832	3	US-08-416-603-1	Sequence 1, Appl
30	25.8	8.6	3881	1	US-08-299-953-2	Sequence 2, Appl
31	25.8	8.6	3881	2	US-08-459-415-2	Sequence 2, Appl
32	25.8	8.6	3881	6	PCT-US95-11231-2	Sequence 2, Appl
33	25.8	8.6	8133	2	US-08-480-604A-5	Sequence 5, Appl
34	25.8	8.6	8133	4	US-08-405-496A-5	Sequence 5, Appl
35	25.8	8.6	51952	5	US-08-947-823-1	Sequence 33, Appl
36	25.4	8.5	878	3	US-08-446-345-33	Sequence 3, Appl
37	25.4	8.5	3825	5	US-08-904-263A-3	Sequence 35, Appl
38	25.4	8.5	6831	4	US-08-446-345-35	Sequence 27, Appl
39	25.4	8.5	10299	4	US-08-609-049A-27	Sequence 1, Appl
40	25.4	8.5	10299	4	US-08-477-451-1	Sequence 5, Appl
41	25.4	8.5	10299	4	US-08-477-451-5	Sequence 25, Appl
42	25.2	8.4	1201	5	US-09-048-889-8	Sequence 8, Appl
43	25.2	8.4	2397	1	US-07-891-942G-11	Sequence 11, Appl
44	25.2	8.4	3807	2	US-08-357-598-1	Sequence 1, Appl
45	25.2	8.4	3807	2	US-08-357-598-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-859-167-1  
; Sequence 1, Application US/08859167  
; Patent No. 6037461  
; GENERAL INFORMATION:  
; APPLICANT: Alnemri, Emad S.  
; APPLICANT: Fernandez-Alnemri, Teresa  
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF  
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS  
; TITLE OF INVENTION: OF MAKING THE SAME  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6037461ris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS  
; SOFTWARE: WordPerfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/859,167  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: TJU-  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1750 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 413..1750  
; US-08-859-167-1

Query Match 9.9%; Score 29.8; DB 5; Length 1750;  
Best Local Similarity 51.1%; Pred. No. 0.85;

Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 54 ggaactgaactgaactactctgttaattgtcatttctaagagcggagcacacctgaa 113  
 Db 365 GGAGCTGTACTGCAAGACCCCTGTGAGCTTCCCTAGTCTAAGAGTAGGATGCTGCTGAA 424  
 QY 114 acccgcaatggtctactagaatcacagaaccccttcagagatgcggggctgcaattctcg 173  
 Db 425 GTCATCCATCAGGTTGAAGAGCACTTGTATACAGATGAGAAGGAGATGCTGCTCTTTTG 484  
 QY 174 aaacaggggtgtgcaat 190  
 Db 485 TGCCGGGATGTTGCTAT 501

RESULT 2

US-09-109-273-1  
 ; Sequence 1, Application US/09109273  
 ; Patent No. 6063760  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alnemri, Emad S.  
 ; APPLICANT: Fernandez-Alnemri, Teresa  
 ; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF  
 ; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6063760ris  
 ; STREET: One Liberty Place, 46th floor  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: WINDOWS  
 ; SOFTWARE: WordPerfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/109,273  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/859,167  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Deluca, Mark  
 ; REGISTRATION NUMBER: 33,229  
 ; REFERENCE/DOCKET NUMBER: TJU-  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 568-3100  
 ; TELEFAX: (215) 568-3439  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1750 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: both  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 413..1750  
 ; US-09-109-273-1

Query Match 9.9%; Score 29.8; DB 5; Length 1750;  
 Best Local Similarity 51.1%; Pred. No. 0.85;  
 Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 54 ggaactgaactggaactactctgttaattgtcatttctaagagcggagcacacctgaa 113  
 Db 365 GGAGCTGTACTGCAAGACCCCTGTGAGCTTCCCTAGTCTAAGAGTAGGATGCTGCTGAA 424

QY 114 acccgcaatggtctactagaatcacagaaccccttcagagatgcggggctgcaattctcg 173  
 Db 425 GTCATCCATCAGGTTGAAGAGCACTTGTATACAGATGAGAAGGAGATGCTGCTCTTTTG 484  
 QY 174 aaacaggggtgtgcaat 190  
 Db 485 TGCCGGGATGTTGCTAT 501

RESULT 3

US-08-484-493-1  
 ; Sequence 1, Application US/08484493  
 ; Patent No. 5728381  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wilson, Peter J  
 ; APPLICANT: Morris, Charles P  
 ; APPLICANT: Anson, Donald S  
 ; APPLICANT: Occhiodoro, Teresa  
 ; APPLICANT: Bielicki, Julie  
 ; APPLICANT: Clements, Peter R  
 ; APPLICANT: Hopwood, John J  
 ; TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF  
 ; TITLE OF INVENTION: IDURONATE 2-SULFATASE  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Scully, Scott, Murphy & Presser  
 ; STREET: 400 Garden City Plaza  
 ; CITY: Garden City  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 11530  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/484,493  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 991,973  
 ; FILING DATE: 17-DEC-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Digiglio, Frank S  
 ; REGISTRATION NUMBER: 31,346  
 ; REFERENCE/DOCKET NUMBER: 84162  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 516-742-4343  
 ; TELEFAX: 516-742-4366  
 ; TELEX: 230 901 SANS UR  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2297 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 125..1774  
 ; US-08-484-493-1

Query Match 9.7%; Score 29; DB 2; Length 2297;  
 Best Local Similarity 53.0%; Pred. No. 1.8;  
 Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 121 atggttactagaagtagcagaagccttcagagatgcggggctgcaattctgcaacagg 180  
 Db 1505 ATTGCTATAGCCAGATATCCCGGCTTCAGACATCCCTCAGTGAATTTCTGACAGCG 1564  
 QY 181 gtgttgactactcaagaaaaattctctgttgataaacactgctagataagaggat 237

100

APPLICANT: Morris, Charles P  
APPLICANT: Anson, Donald S  
APPLICANT: Occhiodoro, Teresa  
APPLICANT: Bielicki, Julie  
APPLICANT: Clements, Peter R  
APPLICANT: Hopwood, John J  
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF  
IDURONATE 2-SULFATASE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/484,493  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 991,973  
FILING DATE: 17-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8416Z  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4428 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 332..434  
FEATURE:  
NAME/KEY: intron  
LOCATION: 536..537  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 693..829  
FEATURE:  
NAME/KEY: intron  
LOCATION: 962..963  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1044..1221  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1350..1351  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1480..1569  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1716..1717  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1841..2041  
FEATURE:  
NAME/KEY: intron  
LOCATION: 2206..2207

FEATURE:  
NAME/KEY: CDS  
LOCATION: 2294..2464  
FEATURE:  
NAME/KEY: intron  
LOCATION: 2585..2586  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2684..2810  
FEATURE:  
NAME/KEY: intron  
LOCATION: 2904..2905  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3033..3206  
FEATURE:  
NAME/KEY: intron  
LOCATION: 3308..3309  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3435..3908  
US-08-484-493-6

Query Match 9.7%; Score 29; DB 2; Length 4428;

Best Local Similarity 53.0%; Pred. No. 2.5;

Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 121 atggtctactagaagtacagaagcccttcagagatgcggggctgcaattctcgaacacagg 180

Db 3636 ATTGCCTATAGCCAGTATCCCGCGCTTCAGACATCCCTCAGTGGAAATTCGACAGCG 3695

QY 181 gtgtgcaattactcaagaaattctctgttgataacactgctagaatagaggat 237

Db 3696 AGTTTAAAGATATAAAGATCATGGGCTATTCCATACGACCATAGACTATAGGTAT 3752

# RESULT 7

US-08-484-494-6  
Sequence 6, Application US/08484494

Patent No. 5798239

GENERAL INFORMATION:

APPLICANT: Wilson, Peter J

APPLICANT: Morris, Charles P

APPLICANT: Anson, Donald S

APPLICANT: Occhiodoro, Teresa

APPLICANT: Bielicki, Julie

APPLICANT: Clements, Peter R

APPLICANT: Hopwood, John J

TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF

IDURONATE 2-SULFATASE

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Scully, Scott, Murphy & Presser

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: New York

COUNTRY: USA

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,494

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 991,973

FILING DATE: 17-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Digiglio, Frank S









; Sequence 1, Application US/08442248  
; Patent No. 5759863  
; GENERAL INFORMATION:  
; APPLICANT: Caras, Ingrid W.  
; APPLICANT: Winslow, John W.  
; TITLE OF INVENTION: AL-1 Neurotrophic Factor  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/442,248  
; FILING DATE: 15-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/330128  
; FILING DATE: 27-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: 920C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-8674  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4165 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-442-248-1

Query Match 9.2%; Score 27.6; DB 2; Length 4165;  
Best Local Similarity 60.8%; Pred. No. 7.5;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 155 tgcggggctgaattctcgaacagggtgttgcaattactcaagaaattctctgttga 214  
Db 988 TCGAAGTCCCCAGTCTCCAGGAAGGCTGTGCAATCCCTCAGAGTAAACTTGAGTTCAA 929  
QY 215 taacactgctagaa 228  
Db 928 TAAAAATTCGGAA 915  
RESULT 14  
US-08-815-1/c  
; Sequence 1, Application US/08440815  
; Patent No. 5798448  
; GENERAL INFORMATION:  
; APPLICANT: Caras, Ingrid W.  
; APPLICANT: Winslow, John W.  
; TITLE OF INVENTION: AL-1 Neurotrophic Factor  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,815  
; FILING DATE: 15-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/330128  
; FILING DATE: 27-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: 920C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-8674  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4165 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-440-815-1

Query Match 9.2%; Score 27.6; DB 2; Length 4165;  
Best Local Similarity 60.8%; Pred. No. 7.5;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 155 tgcggggctgaattctcgaacagggtgttgcaattactcaagaaattctctgttga 214  
Db 988 TCGAAGTCCCCAGTCTCCAGGAAGGCTGTGCAATCCCTCAGAGTAAACTTGAGTTCAA 929  
QY 215 taacactgctagaa 228  
Db 928 TAAAAATTCGGAA 915

RESULT 15  
US-08-618-464-1/c  
; Sequence 1, Application US/08618464  
; Patent No. 5773570  
; GENERAL INFORMATION:  
; APPLICANT: CARSON, DENNIS A.  
; APPLICANT: ALBANI, SALVATORE  
; TITLE OF INVENTION: VACCINE COMPOSITIONS AND METHODS USEFUL IN  
; IMMUNE PROTECTION AGAINST ARTHRITIS  
; TITLE OF INVENTION: INDUCING  
; TITLE OF INVENTION: INVOLVED IN  
; TITLE OF INVENTION: THE PATHOGENESIS OF RHEUMATOID ARTHRITIS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & RICHARDSON P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CALIFORNIA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/618,464  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HOWELLS, STACY L.  
; REGISTRATION NUMBER: 34,842  
; REFERENCE/DOCKET NUMBER: 07340/042001

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1055 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (clonal)
; IMMEDIATE SOURCE:
; CLONE: E. coli K12 dhaj polynucleotide (cdna)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1054
; US-08-618-464-1
```

```
Query Match 9.1%; Score 27.4; DB 2; Length 1055;
Best Local Similarity 62.3%; Pred. No. 4.4;
Matches 43; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 150 agagatgcggggctgcaattctcgaaacagggtgttgcaattactcaagaaaaattctctg 209
   || || | | | | | | | | | | | | | | | | | | | |
Db 740 AGCGAAGTTGATCGGGAATTCGCAATACAGTTTGTTCCTTCACGCTCGAAAAATCGGGTG 681

QY 210 ttggataac 218
   || || | | |
Db 680 CTGTTTAAAC 672
```

Search completed: November 4, 2000, 13:34:28  
Job time: 16704 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:37:55 ; Search time 4075.18 Seconds  
(without alignments)  
412.624 Million cell updates/sec

Title: us-09-300-482-356

Perfect score: 385

Sequence: 1 caaccggcaccagcctgac.....acggtctgggagttttttt 385

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_bal:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pil:\*
- 8: gb\_pl2:\*
- 9: gb\_prl:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: em\_fun:\*
- 13: em\_hum1:\*
- 14: em\_hum2:\*
- 15: em\_in:\*
- 16: em\_om:\*
- 17: em\_or:\*
- 18: em\_ov:\*
- 19: em\_pat:\*
- 20: em\_ph:\*
- 21: em\_pl:\*
- 22: em\_ro:\*
- 23: em\_sts:\*
- 24: em\_sy:\*
- 25: em\_un:\*
- 26: em\_vi:\*
- 27: gb\_htg1:\*
- 28: gb\_htg2:\*
- 29: gb\_in1:\*
- 30: gb\_in2:\*
- 31: em\_ba1:\*
- 32: em\_ba2:\*
- 33: em\_hum3:\*
- 34: em\_hum4:\*
- 35: gb\_pr4:\*
- 36: gb\_htg3:\*
- 37: gb\_htg4:\*
- 38: gb\_htg5:\*
- 39: gb\_htg6:\*
- 40: gb\_htg7:\*
- 41: em\_htg1:\*
- 42: em\_htg2:\*
- 43: em\_htg3:\*

44: em\_hum5:\*

45: gb\_pl3:\*

46: gb\_pr5:\*

47: gb\_htg8:\*

48: gb\_htg9:\*

49: gb\_htg10:\*

50: gb\_htg11:\*

51: gb\_htg12:\*

52: gb\_htg13:\*

53: gb\_htg14:\*

54: gb\_in3:\*

55: gb\_htg15:\*

56: gb\_htg16:\*

57: gb\_htg17:\*

58: em\_htg4:\*

59: em\_htg5:\*

60: em\_htg6:\*

61: em\_htg7:\*

62: em\_hum6:\*

63: gb\_htg18:\*

64: gb\_htg19:\*

65: gb\_ba3:\*

66: em\_htg8:\*

67: em\_htg9:\*

68: em\_htg10:\*

69: em\_htg11:\*

70: em\_htg12:\*

71: em\_htg13:\*

72: em\_htg14:\*

73: em\_htg15:\*

74: em\_htg16:\*

75: em\_htg17:\*

76: em\_htg18:\*

77: em\_htg19:\*

78: em\_htg20:\*

79: em\_htg21:\*

80: em\_htg22:\*

81: em\_htg23:\*

82: gb\_pr6:\*

83: gb\_pr7:\*

84: gb\_htg20:\*

85: gb\_htg21:\*

86: gb\_htg22:\*

87: gb\_htg23:\*

88: gb\_ro:\*

89: gb\_sts1:\*

90: gb\_sts2:\*

91: gb\_sy:\*

92: gb\_un:\*

93: gb\_vil:\*

94: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	% Match	Query Length	DB ID	Description
C	1	247.8	64.4	101441	8	AF002542 Oryza sat
	2	187.8	48.8	2319	8	Y15781 Capsicum an
	3	177.8	46.2	1802	8	Z46646 C.plantagin
	4	170.4	44.3	2629	5	A52295 Sequence 1
	5	170.4	44.3	2629	5	AR072009 Sequence
	6	170.4	44.3	2629	5	EL3696 cDNA encodi
	7	167	43.4	2447	45	Z50099 S.tuberosum
	8	165.4	43.0	2232	8	Z46647 C.plantagin
	9	163.8	42.5	2207	8	Z46647 C.plantagin
	10	162.4	42.2	2519	45	L76554 Spinacia ol
	11	162.2	42.1	107865	8	AL162295 Arabidops
	12	155.6	40.4	2595	8	AJ249787 Cyanophor





[illegible]





ORGANISM

Craterostigma plantagineum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophytes; Magnoliophyta; eudicotyledons; Asteridae; Gentiananae; Lamiales; Scrophulariaceae; Craterostigma.

REFERENCE

1 (bases 1 to 2232)

REFERENCE	1 (bases 1 to 2232)
AUTHORS	Bartels, D.
TITLE	Direct Submision
JOURNAL	Submitted (10-NOV-1994) Bartels D., Max-Planck-Institut fuer Zuechtungsforschung, Pflanzenzuechtung und Ertragsphysiologie, Carl-von-Linne-Weg 10, Cologne, NRW, Germany, 50829
REFERENCE	2 (bases 1 to 2232)
AUTHORS	Bernardini, G., Schwall, G., Lottspeich, F., Salami, F. and

**TITLE** Bartels, D.  
The transketolase gene family of the resurrection plant  
*Cratogeomys plantaginifolius*: differential expression during the  
rehydration phase

**JOURNAL** EMBO J. 14 (3), 610-618 (1995)

**MEDLINE** 95163594

FEATURES	source	Location/Qualifiers
		1..2232
		/organism="Craterostigma plantagineum"
		/db_xref="taxon:4153"
		/clone="Cpl-tkt7"
gene		29..2059
		/gene="Cpl-tkt7"
CDS		29..2059

```

/gene="Cpl-1ktk7"
/EC_number="2.2.1.1"
/citation={2}
/codon_start=1
/product="transketolase"
/protein_id="CAA86609.1"
/db_xref="GI:664903"
/db_xref="SWISS-PROT:O42677"

```

translation = MAPK111APRFLVSKSVNTIRFLADAVKAKSGHPMPKCMG  
PMGHVLYDEEWRENPKNPYENRDFVFLSAGHCMLQYALLHSLGYSDEEDKLKSL  
WMSSTPTPAHPENFTPGVETVTPGLGGLIASALVAAEKHLAAAYNKPQFVLDH  
VYLLGDGQCMFEGVSENAALHNGKGLIALDDNHLITDGTDTVAEDVDKRFRT  
LGWHVYKNGNDSCDELRALAEKSVKDRPPIKVTTTIGYGAQSKANTYGVHG  
LGPKEAETRNKLGWYEPFHVDDKVKHHSRHLABGAALAEAPAKFAQKFFP  
ADLAKSITGELGSPNIPYTPYTPENPGLPTRLTSLQILINGLDGVLPLGLGGSAD  
JLNNMFLKNGDGFQKSGEIRNVFGEAREHMGSGICNGLAHLGGLLPPCYATVFV  
I

BASE COUNT  
ORIGIN

	Query Match	43.0%	Score 165.4;	DB 8;	Length 2232;
	Best Local Similarity	67.6%;	Pred No. 1.3e-21;		
	Matches 222;	Conservative	Mismatches 111;	Indels 0;	Gaps
QY	12	aagcctgacatcatattgggttgccaccgcctccgagctcgtagatcgcgggcgaatcgccc	71		
Dd	1712	AAGCCCGACGCTCACTACTGATCGGGATCGGCTCGAGCTTGAGATAGCGGTAGGCCCGG	1771		

Qy	72	gacgactgaggaagagaggaagacggcccgagtggtctctgattggtctctcggaactc	131
Db	1772	GACGAGCTCAGGAGGAGGAGGCAAAAGTCCGATGTTCTCGCTTGTGTGTGGCAGTTG	1831
Qy	132	tttgatgagcagtcggatagtcacaaagagagagcgtccctccctcgagaggtcacacgagag	191
Db	1832	TTCCCGGACGACGTCGCGAGAGTATCAGAGACGGTTTTCCTCCGCGGGGTTCAGCGCTAGG	1891

Qy	192	atcagcatcagagccggggtcccaactcggctgcgagaaagtaagtcgagaccgaagcaag	251
Db	1892	GTTAGCTGAGCGGGCCGCGCTCGACTTTCGGTGGGAGGGTTTATCGGGCCGAGGGTAA	1951
Qy	252	gccattgcatacacaagttcgcgcgagtagtctcctgcgggacagatctacaaagagatc	311
Db	1952	CGGTTGGGATCATAGTTTGGGGCGACCCGCCGGGAGAGAGCTTTTAAAGAGTTT	2011

•

Qy	312	ggcatcac	cgtagag	agagcat	cattgca	actgcc	aagagct	ttt	354
Db	2012	GGGATA	ACCGT	CGAAG	CACTG	TCGCG	GGCGA	AGGATTT	2054

RESULT	9	
CPTKT10		
LOCUS	2207 bp	mRNA
DEFINITION	C.plantagineum tkt10 gene for transketolase.	
ACCESSION	246647	
VERSION	246647.1	GI:664898

**KEYWORDS** tkt10 gene; transketolase.

**SOURCE** Craterostigma plantagineum.

**ORGANISM** Craterostigma plantagineum  
Eukaryota; Viridiplantae; Streptophyta; Embryophytes; Magnoliophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Asterridae; Gentiananae; Lamiales; Scrophularia

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 (bases 1 to 2207)  
Bartels, D.  
Direct Submission  
Submitted (10-NOV-1994) Bartels D., Max-Planck  
zuechtungs-forschung, Pflanzenzuechtung und Ertr-  
Carl-von-linne-Weg 10, Cologne, NRW, Germany,  
2 (bases 1 to 2207)  
Bernachia, G., Schwall, G., Lottspeich, F., Sala  
REFERENCE  
AUTHORS

TITLE	Barbels.D. The transketolase gene family of the resurrected <i>Cratogeomys plantagineum</i> : differential expression during rehydration phase
JOURNAL	EMBO J. 14 (3), 610-618 (1995)
MEDLINE	95163594
FEATURES	Location/Qualifiers 1. .2207 source

```

gene
CDS
    /organism="Craterostigma plantaginineum"
    /db_xref="taxon:4153"
    /clone="Cpl-tkl10"
    3..2042
    /gene="Cpl-tkl10"
CDS
    3..2042
    /gene="Cpl-tkl10"
    /EC_number="2.2.1.1"

```

```

/citation={2]
/codon_start=1
/product="transketolase"
/protein_id="CAA86608.1"
/db_xref="GI: 664899"
/db_xref="SWISS-PROT:Q42675"
/translation="MAKTPSPSAAALAEIVKVSVA
MGCAPMGHVLDFEKNPKNPYFNEDRFVLSAGH

```

KGLRWGSKFTPAHNPETPGVEVTTGTPLGQGVGSSA  
 DHYVTIMGLDGGCMGEGISNEASSLAHHWGLCKLIAL  
 RFEALGWHVTVVANGNDGQDEIRAIKQVAKSVTDKPT  
 HNALPGAEARTNQLGWCPEYTFHFVDDVKVKKHWSIR  
 YPEKAAELKSLITGELGWKELAPTYTPSPGNPTPL  
 ADLTASNMAFLKSSGDFQKPTGRNLKFGAREHGMC  
 YFTYMRRAAIVRAALSKARVVIMTHDSIGLGDGP  
 ADGNETAGAYKVAANAGRPSILSLRQKLPQLPGT

BASE COUNT	520 a	582 c	682 g	423 t
ORIGIN				
Query Match		42.5 %	Score 163.8;	DB 8;
Best local similarity		68.1 %	Pred. No. 2, 6e-21;	

	Matches	228;	Conservative	0;	Mismatches	107;	Indels
QY	12	aagcctgacatcattgggttggtggccagcgcctccgagctggagatccgc					
Db	1695	AAGCCGAGGTGATACACTGCGGACGGCTCGGAGCTTGAGATAGCC					
QY	72	gacgagctgaggaagaggagggaagacggtccgcgctgcttcgttcgtt					

---







**LOCUS** AC002387 122871 bp DNA PLN 05-APR-2000  
**DEFINITION** Arabidopsis thaliana chromosome II section 242 of 255 of the complete sequence. Sequence from clones T14P1, F4L23.  
**ACCESSION** AC002387 AE002093  
**VERSION** AC002387.2 GI:6598365  
**KEYWORDS** HFG.  
**SOURCE** thale cress.  
**ORGANISM** Arabidopsis thaliana  
**REFERENCE** Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
**AUTHORS** 1. (bases 1 to 122871)  
 Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanKen,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,W.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.  
**TITLE** Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana  
**JOURNAL** Nature 402 (6763), 761-768 (1999)  
**MEDLINE** 20083487  
**REFERENCE** 2 (bases 1 to 122871)  
**AUTHORS** Lin,X.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
**COMMENT** On Dec 17, 1999 this sequence version replaced gi:2583106. The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site <http://www.tigr.org/tdb/at/at.html>.  
 Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.org/pub/pub/xgrail>), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/GENSCANW.html>), and NetPlantGene (<http://www.cds.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/repeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.  
 We thank the GSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T1A5, and T13L16, the ESSA group for sequencing clones F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.  
 This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.  
 Address all correspondence to: [at@tigr.org](mailto:at@tigr.org).  
**FEATURES** Location/Qualifiers  
**source** 1..122871  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="II"  
 complement(<1..2952)  
 /note="Sequence from clone T14P1"  
 /rpt\_family="(TAAA)n"  
 join(2227..2322,2480..2527,2650..2705,3098..3150,3250..3641)  
 /gene="At2g45170"  
 2227..3641  
 /gene="At2g45170"  
 /note="T14P1.2"  
 join(2488..2527,2650..2705,3098..3150,3250..3469)  
 /gene="At2g45170"  
 /codon\_start=1  
 /product="putative microtubule-associated protein"  
 /protein\_id="AAF18602.1"  
 /db\_xref="GI:6598366"  
 /translation="MNRGSIFKMDDDFEKKRAEAGRIREKYPDRIPIVIVEKAEKSEVP  
 NIDKKYLVPSDLTVGQFVVYIRKRIKLSAEKALIFVDNVLPTGELMSSVYEDKKD  
 EDGLYITYSGENTEGASSI"  
 complement(2953..>122871)  
 /note="Sequence from clone F4L23"  
 4282..4870  
 /gene="At2g45180"  
 4282..4870  
 /gene="At2g45180"  
 /note="F4L23.31"  
 4320..4724  
 /gene="At2g45180"  
 /codon\_start=1  
 /product="putative proline-rich protein"  
 /protein\_id="AAB82643.1"  
 /db\_xref="GI:2583134"  
 /translation="MASKALAVTALLITLNLFFFTVSTKCPPTTKPKPKTKSPKK  
 APAVKPTCDTTLKGLGVNVMVVGSPKPTCCCTLLQGLANLEAAVCLCTALK  
 ANVILGNVPIIDLTLNLYCGKRVPHGQCS"  
 5379..5403  
 /rpt\_family="POLY A"  
 complement(join(<6393..6470,6586..6663,7082..7157,7248..7296,7872..8004,8099..8260,8382..8495))  
 /gene="At2g45190"  
 complement(<6393..>8495)  
 /gene="At2g45190"  
 /note="F4L23.30"  
 complement(join(6393..6470,6586..6663,7082..7157,7248..7296,7872..8004,8099..8260,8382..8495))  
 /gene="At2g45190"  
 /note="unknown protein"  
 /codon\_start=1  
 /protein\_id="AAB82644.1"  
 /db\_xref="GI:2583135"  
 /translation="NMLSSMSVSSAVCSVDHPSDHLCYVQCNCQITLAVNVPTT  
 SLKFTVTRCGCTTLLSVNMRSYVLPSANQLQLGPHSYFNPDIIELRDAPSNM  
 NMMNMGHTPMNDIPFMDLHQOHEIPKAPVNRPEKRVPSAYNRFIKERIQRIK  
 AGNPDISHREAFSAAKANWAHFPHIHFGLVPDNPVKVTNMQEGEDNMVKEGFA  
 PAANVGVTPT"  
 10147..10201  
 /rpt\_family="(GA)n"  
 12613..12646  
 /rpt\_family="(CATA)n"  
 complement(join(15613..15729,15821..15950,16066..16137,16252..16341,16810..17006,17451..17647))  
 /gene="At2g45200"  
 complement(15613..17647)  
 /gene="At2g45200"  
 /note="F4L23.29"  
 complement(join(15632..15729,15821..15950,16066..16137,16252..16341,16810..17006,17451..17583))  
 /gene="At2g45200"  
 /codon\_start=1  
 /product="putative cis-Golgi SNARE protein"  
 /protein\_id="AAB82642.1"

```

/db_xref="GI:2583133"
/translation="WTESSLDLQESWEELREARKIEGDLVDKLVSKYAKIGARFTQG
GYVFGSTVSGRSWKSMEIEIQSLLEKLDINDMSRCAASAPPTSVYQKLARHR
DILYTFEERIKRINSLREHAELLSVRDDI SEYKASGSMSPGVVLRERASIHG
SLSHDDVIGQAQTRAVIGSQSLFSDVOGVKVLGDKFPVIRGLLGSIKRKRSDT
LILSLAIACTFLIILYWSK"
misc_feature
complement(17986..18862)
repeat_region
/note="molecular marker ML"
complement(18449..18564)
/rpt_family="(PAAAA)n"
19741..>20315
/gene="At2g45210"
19741..>20315
/gene="At2g45210"
/feature="F4L23.28; member of auxin-induced protein multigene
family"
19827..>20315
/gene="At2g45210"
/codon_start=1
/product="putative auxin-induced protein"
/protein_id="BAB82641.1"
/db_xref="GI:2583132"
/translation="MRKIIGRIGRRVSWIFRKTIRQSYNRHSTQOACMLMRPL
AKLSWGRLQSRFRRSRAVPIVDHKKADVPVRGHAIYVVGQDGDCHRVLPV
VYFNHPLFGLLEAREAEKEYGFCHEGGITPCLYSDFERVKTRIASGSSRVFPWGRHC
RN"
misc_feature
complement(join(<22224..22909,23488..>24334))
19827..>20315
/gene="At2g45220"
complement(<22224..>24334)
/gene="At2g45220"
/feature="F4L23.27"
complement(join(22224..22909,23488..>24334))
/gene="At2g45220"
/codon_start=1
/product="putative pectinesterase"
/protein_id="BAB82640.1"
/db_xref="GI:2583131"
/translation="MAPRAYTINFEVILVASTVSGYNOKVAKWCSOTPNPKPCBY
FLTHSNNEPIKSESEELKISMKVLDRAILAKTHATLPGKRDTRKRAWEDCIKL
YDLTVSKNETMDPNKSKLDAQTWLTALNTDTRAGFLGVDIVLPLMSNVV
SNLCLNTAIKNKPNFTPEKDGFPKSWKPKQLQSSPTPKDNVAVKDGSGNFT
IKEADAASGSGREVIYVKGQVSEENLEIRKKNVLRGDKIGKTITGSKSVGGTTT
FNSATVAAGVDGFTGRTFRNTAGASNOEAVALRSGDLSVYQCSFEAVODTLVYH
SNQFYRDCDVGTVDFTFGNAAVLONCNIFARRPSKNTLITAGRSDPNQNTGII
IHNSRVTAAADLRVLGSKTYLGRPHQVSRVTFMTKSLDLSLDIPRGWLENDGFAL
KILYFAEQNTGPGASTSGRVTFGRVLSASEASKFTVTGTFILAGGWSIPVFTS
GL"
repeat_region
23009..23038
/rpt_family="(TA)n"
repeat_region
28701..28741
/rpt_family="(POLY_A"
repeat_region
complement(29132..29228)
/rpt_family="(POLY_A"
misc_feature
complement(<29263..>33387)
Query Match
40.3%; Score 155; DB 7; Length 122871;
Best Local Similarity 65.4%; Pred. No. 3.7e-20;
Matches 227; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 1 caacggcaccagcgtgacatcattgggttggccaccggtccgagctggagatcgcg 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53174 CAACCGGTAAACACCGATGTGATCTTGTCGGAACCTGGATCAGAGCTAGATGCTG 53233
QY 61 gcaatcgccgacagctgaggaaggagggaagacgggtcgogtctctctgctgtc 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53234 CTCAGAGCTGCAGAGAAGTAAAGGCAACAGGCAAAAGTGTGAGAGTTGTTCCCTTTGAT 53293
QY 121 cctgggaactctttgatgacagctcgatgacacagagagcgtctctccctccgacg 180
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53294 GTTGGAGCTTTTCACGAGCAATCAGATGCATACGAAGAAGTGTGTCACATCTGATG 53353
QY 181 tcacagcaggaatcagcagcagcggggtccactctcgctcggtggcagaagtcgag 240
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 53354 TATCAGCTAGAGTTAGTATCGAAGTCGATCGACTTTTGGATGGGAAAGATCTCGAG 53413
QY 241 cccaagcaagccattggcattgcacgaattcgcgcgagtgctcttcgcgggacgatct 300
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53414 GAAAAGGGAATCGATTGGAATTGATAGTTTGGAGCAAGTGCACACGAGGAAAGCTTT 53473
QY 301 acaaggagtaagcattaccctggagagcagcattgcaactgccaag 347
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53474 ATAAAGAGTTTGGTATCACCATTGAAGCTATGTTGAAGCAGCAAG 53520
RESULT 14
F6N15/c
LOCUS F6N15 95643 bp DNA PLN 09-JUN-1998
DEFINITION Arabidopsis thaliana BAC F6N15.
ACCESSION AF069299
VERSION AF069299.1 GI:3193311
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 95643)
Washington University Genome Sequencing Center.
The A. thaliana Genome Sequencing Project
Unpublished (1997)
REFERENCE
2 (bases 1 to 95643)
Ryan, E., Edwards, J. and Pape, K.
The sequence of A. thaliana F6N15
Unpublished (1998)
JOURNAL
REFERENCE
3 (bases 1 to 95643)
Waterston, R.
Direct Submission
Submitted (29-MAY-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu

```

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

#### NEIGHBORING COSMID INFORMATION:

The 3' cosmid is F5110, 1300 bp overlap. Actual start of this cosmid is at base position 1 of F6N15; actual end is at 1950 F5110.

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).  
Location/Qualifiers  
1. .95643  
/organism="Arabidopsis thaliana"

FEATURES  
source







introns  
complement(6081..6143)  
/number=1  
complement(6144..6224)  
/gene="AT4g00020"  
/number=2  
complement(6225..6293)  
/number=2  
complement(6294..6375)  
/gene="AT4g00020"  
/number=3  
complement(6376..6452)  
/number=3  
complement(6453..6638)  
/gene="AT4g00020"  
/number=4  
complement(6639..6711)  
/number=4  
complement(6712..6813)  
/gene="AT4g00020"  
/number=5  
complement(6814..6913)  
/number=5  
complement(6914..6947)  
/gene="AT4g00020"  
/number=6  
complement(6948..7000)  
/number=6  
complement(7001..7139)  
/gene="AT4g00020"  
/number=7  
complement(7140..7243)  
/number=7  
complement(7244..7368)  
/gene="AT4g00020"  
/number=8  
complement(7369..7462)  
/number=8  
complement(7463..8437)  
/gene="AT4g00020"  
/number=9  
complement(8438..8685)  
/number=9  
complement(8686..8847)  
/gene="AT4g00020"  
/number=10  
complement(8848..8981)  
/number=10  
complement(8982..9125)  
/gene="AT4g00020"  
/number=11  
complement(9126..9210)  
/number=11  
complement(9211..9364)  
/gene="AT4g00020"  
/number=12  
12565..13366  
/gene="AT4g00030"  
12565..13028  
/gene="AT4g00030"  
/number=1  
join(12565..13028,13114..13179,13258..13366)  
/gene="AT4g00030"  
/notes=coded for by A. thaliana cdna W43412  
contains EST gb:W43412  
/codon\_start=1  
/product="predicted protein of unknown function"  
/protein\_id="CAB80761.1"  
/db\_xref="GI:7267090"  
/translation="MALALSLSACSPPLRRTRAGPRTSCSIFANPAORAKRKLLELI  
SEDRGLRTQDKPKRDEIVNAESMTVIGRSSITDDSDLSATWRLTWTKEOLFII  
EKAGLFGTQAGDVLYDYNKRILNNVITFPDGVFVRSDDIDIASPQVNFENSAV  
LRGNWELPLPPFGKGFENVYMDGEIRVAKDRIGDYLIVDRAPYNWTSFV"

introns  
13029..13113  
/gene="AT4g00030"  
/number=1  
13114..13179  
/gene="AT4g00030"  
/number=2  
13180..13257  
Query Match 35.2%; Score 135.6; DB 8; Length 197119;  
Best Local Similarity 63.9%; Pred. No. 1.3e-16;  
Matches 221; Conservative 0; Mismatches 124; Indels 1; Gaps 1;  
QY 9 accaagcctgacatcattggttgggacccggctccgagctggagatcgcgggcaatgcg 68  
Db 40211 ACAAACTAGACGTAATCCTAATGGGGACAGGTCGGAGCTAGAGATAGCTGCAAGGCT 40152  
QY 69 gccacgagctgaggaaggggaagacggctccgagctgcgtcgtctcgttcgtctcctgggaa 128  
Db 40151 GGAGAGAAAGCTAAGAGAGAAAGAAAGACAGTGAAGTGTGTCATTGGTGAGTTGGGAG 40092  
QY 129 ctcttgatgagcagtcggtatgatacaagagagagcgtccctccctgcccagctcacagcg 188  
Db 40091 CTGTTTGATGAACAAGAGAGAGAAATACAAAGAGTGTGTACCTTGAGAGGTTTCAGCA 40032  
QY 189 aggatcagcatcgagccgggtccactctcgctgggagagagtagctcggtggagcccaagggc 248  
Db 40031 AGACTGAGCATTGAAGCAGGTCGACCTTTGGTGGGAGAGATGGTGGCTCCCAAGA 39972  
QY 249 aaggccattggcatcgacaaagttcggcgagtgctcctcgcgggagacgatacagaagag 308  
Db 39971 AAAGCCATCGGTGTAGAC-AGTTTGGAGCAAGTGCACGCAATGTGCTCTACAGGAAA 39913  
QY 309 tacggcatcacccgtgagagcatcattgcaactgccaagagcgtttt 354  
Db 39912 TTGGACTCACCGTGGATACGGTTATTGACGCCGCTAAGACTTTT 39867

Search completed: November 4, 2000, 13:39:32  
Job time: 17421 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:31 ; Search time 320.8 Seconds  
(without alignments)  
450.842 Million cell updates/sec

Title: US-09-300-482-356

Perfect score: 385

Sequence: 1 caaccggcaccagcctgac.....acggtctggagttttttt 385

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	170.4	44.3	2629	17	T35903	Sequence encoding
2	80.6	20.9	11443	19	V52182	Streptococcus pneu
3	80	20.8	1953	18	V74685	Staphylococcus aur
4	63	16.4	1738	19	T98569	DNA encoding a S.
5	55.8	14.5	3225	20	X13233	Enterococcus faeca
6	53.8	14.0	946	19	V52372	Streptococcus pneu
7	51.6	13.4	1872	17	T28491	P. aeruginosa dete
8	44	11.4	1284	21	Z46836	Heat-resistant iso
9	44	11.4	1284	21	Z57439	Heat-resistant iso
10	43	11.2	1281	21	Z46835	Heat-resistant iso
11	43	11.2	1281	21	Z57438	Heat-resistant iso
12	43	11.2	1480	19	V35659	Heat-resistant iso

13	43	11.2	2872	21	Z46834	Heat-resistant iso
14	43	11.2	2872	21	Z57437	Heat-resistant iso
15	41.2	10.7	4195	17	T12068	FLT4 receptor tyro
16	41.2	10.7	4425	16	T03090	Protein tyrosine-k
17	41.2	10.7	9108	16	T03104	Plasmid prk5.tki-1
18	41	10.6	1693	21	Z56005	Contig 004 from co
19	39	10.1	1380	20	X22117	Human secreted pro
20	37.2	9.7	723	15	Q73597	Dactylis glomerata
21	36.8	9.6	1008	21	A27338	Rice RA8 anther-sp
22	36.8	9.6	1910	19	V48148	Nicotianamine amin
23	36.8	9.5	3003	21	A27333	Rice RA8 anther-sp
24	36.6	9.5	2109	20	Z07194	Human lung tumour
25	36.2	9.4	5821	18	T58686	DNA encoding S. fr
26	35.6	9.2	1136	16	O89541	Acetyl-CoA-acyltra
27	35.6	9.2	1431	16	O85641	Wheat cold toleran
28	35.6	9.2	1632	14	Q42976	Carminomycin 4-O-m
29	35.6	9.2	4980	14	Q36660	Sequence of the po
30	35.4	9.2	68750	21	Z55887	Sorangium cellulos
31	35	9.1	77536	21	A14651	Nucleotide sequenc
32	34.8	9.0	1026	20	Z28362	Platelet factor-4
33	34.8	9.0	1918	19	V68515	STRL33 coding sequ
34	34.8	9.0	2238	19	V24017	Human HMBU14 codi
35	34.6	9.0	1157	12	Q13716	Geranylgeranyl pyr
36	34.6	9.0	1157	12	Q13717	Geranylgeranyl pyr
37	34.6	9.0	1157	17	T40789	Geranylgeranyl pyr
38	34.6	9.0	1157	17	T40790	Recombinant gerany
39	34.6	9.0	1157	17	T41741	Wild type geranyig
40	34.6	9.0	1157	17	T41742	N-terminally trunc
41	34.6	9.0	1157	17	T37091	Geranylgeranyl pyr
42	34.6	9.0	1157	17	T37092	E. herbicola heran
43	34.6	9.0	1157	18	T91541	Geranylgeranyl pyr
44	34.6	9.0	1157	18	T91542	Erwinia herbicola
45	34.6	9.0	12905	21	Z99101	S. fradiae tylosin

#### ALIGNMENTS

RESULT 1

T35903  
ID T35903 standard; cDNA; 2629 BP.

XX T35903;

XX 29-JAN-1997 (first entry)

XX Sequence encoding tobacco transketolase.

XX Transketolase; herbicide; Nicotiana tabacum; identification; screening; ss.

XX Nicotiana tabacum.

XX Key Location/Qualifiers  
CDS 60..2291

FT /\*tag= a  
FT /product= Transketolase.

XX EP723017-A2.

XX 24-JUL-1996.

XX 13-JAN-1996; 96EP-0100458.

XX 23-JAN-1995; 95DE-4001906.

XX (SCHM/) SCHMIDT R.

XX (BADI) BASF AG.

XX Schmidt R, Sonnwald U, Stitt M;

XX WPI; 1996-335478/34.

DR P-PSDB; W03319.

XX	New trans:ketolase from tobacco and related DNA - useful for screening for inhibitors which are potential herbicides
pt	Claim 5; Page 9-12; 26pp; German.
XX	The transketolase from tobacco is useful for identifying herbicides for their ability to inhibit the transketolase. The transketolase can be produced in pure form by cloning the gene.
CC	Sequence 2629 BP; 686 A; 595 C; 620 G; 728 T; 0 other;
XX	
XX	Query Match 44.3%; Score 170.4; DB 17; Length 2629;
XX	Best Local Similarity 67.4%; Pred. No. 4e-34;
XX	Matches 240; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY	4 ccggcaccaacgctgcacatcattgggttggccaccggctccagctggagatcgcgaggca 63 
Db	1936 ctggacaacacctgatctatttgattggtactggtcagagttagaaattgctgtca 1995
QY	64 atgcgcccagcagctgagaagagggggaagcgtccgctcgctctgcttcctcct 123 
Db	1996 aggcctgctgatgaactcaggaaagaaaacagtcagagttgttctcttgttgtt 2055
QY	124 gggaaacttttgatgagcagtcggatgagtacaaagagagcgtcctccctgccgacgtca 183 
Db	2056 ggagacttttgaagaacaatacagccgactacaaaggaaagtgctctccatcatctgtta 2115
QY	184 cagcagagatcagcatcagccgggtccactctcgctggcagaaagtcgtagagccc 243 
Db	2116 cagctcagagttagcattgagccggatccacatttgggtgggagaaatatgctcgatcaa 2175
QY	244 aaggcaagccattggcatcgacaaagttcggcgagagtcctcctgccgggacgatetaca 303 
Db	2176 aggggaagccatcggaattacagatggggtgccagtgcctctgtggaataatata 2235
QY	304 aggagtagcgcateaccgtggagagcatcattgcaactgccaaagcgttttaagag 359 
Db	2236 aggagtagcgaattacagcagaggtgtgtgctgcagctaacaaagtttcttag 2291
RESULT	2
V52182	
ID	V52182 standard; DNA; 11443 BP.
AC	V52182;
XX	
DT	23-OCT-1998 (first entry)
XX	Streptococcus pneumoniae genome fragment SEQ ID NO:49.
DE	
XX	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW	computer readable medium; vaccine; pharmaceutical composition; ds.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	WO9818931-A2.
XX	
PD	07-MAY-1998.
XX	
PF	30-OCT-1997; 97WO-US19588.
XX	
PR	31-OCT-1996; 96US-0029960.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI	Kunsch CA, Rosen CA;
XX	
DR	WPI; 1998-272225/24.
XX	
PT	Computer-readable medium with recorded Streptococcus pneumoniae

polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae

Claim 1; Page 445-451; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (V52134 to V52524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridise to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the S. pneumoniae genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the S. pneumoniae genome of commercial importance, or expression modulating fragments of the S. pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for S. pneumoniae.

Sequence 11443 BP; 3350 A; 2057 C; 2533 G; 3502 T; 1 other;

Query Match            20.9%; Score 80.6; DB 19; Length 11443;  
Best Local Similarity   51.9%; Pred. No. 1.9e-11;  
Matches 182; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY     17   tgacatcatgttggttggccaccgctcgcagctcgaggatcgcgggcaattcggcgaaca 76  
      ||| | | | | | | | | | | | | | | | | | | | | |  
Db   10003 tgcatacttatcttgtctcacagtctcgaggtccaacttagctatcaaagctgctaaga 10062

QY     77   gctcgaagaaaggagggaagacggtcccgcgtcttcgtcttcctcggsaacctttga 136  
      ||| | | | | | | | | | | | | | | | | | | | | |  
Db   10063 attgttttacaaagtggttaagtaactgtgtgatctatgccccaacsgaactattga 10122

QY     137   ttgacagctcggatgagtacaagagagcgctcctccctgccgcagctcacgcgaggatacg 196  
      ||| | | | | | | | | | | | | | | | | | | | | |  
Db   10123 tgcacaagatgctacctacaagggaagacattttaccatctaaagactcgtcgtgtgac 10182

QY     197   catcgagcgccggtccactctcgctgcgagaagtacgctcgagccccaaaggccaagcccat 256  
      ||| | | | | | | | | | | | | | | | | | | | | |  
Db   10183 cattgaaaatggcagcgaccaaaagttgtgtacaaagtattgtgttgatggcgcggtcat 10242

QY     257   tggcatcgacaaagtctggcgagtgctcctgccgggacgactcataagagagatggcgcatt 316  
      ||| | | | | | | | | | | | | | | | | | | | | |  
Db   10243 cggtattgacatctctggfctgctgccacagctcagactgtgattgataatatgtgatt 10302

QY     317   cacgctggagagcactaatggcaactgccaaagactttttaagagataacaac 367  
      ||| | | | | | | | | | | | | | | | | | | | | |  
Db   10303 tacggttagaagatcatcgtctgcgaagttaaaccctcatagaaacccaattac 10353

RESULT      3

V74685

ID   V74685 standard; DNA; 1953 BP.

XX   V74685;

AC   V74685;

XX   V74685;

XX   V74685;

DT   16-MAR-1999 (first entry)

XX   XX

DE   Staphylococcus aureus contig SEQ ID #374.

XX   Computer readable medium; vaccine; S aureus infection; immunodetection;  
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

Computer readable medium; vaccine; *S. aureus* infection; immunodetection; cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;





XX  
PS  
XX  
Claim 1; Page 1259-1260; 1409pp; English.

CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from *Streptococcus*  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the *S. pneumoniae*  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridise to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the *S. pneumoniae* genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC *S. pneumoniae* genome of commercial importance, or expression modulating  
CC fragments of the *S. pneumoniae* genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for *S. pneumoniae*.  
XX

SQ Sequence 946 BP; 272 A; 189 C; 218 G; 267 T; 0 other;

5Q Sequence 946 BP; 272 A; 189 C; 218 G; 267 T; 0 other;

[illegible]

RESULT	7
T28491	T28491
DD	T28491 standard; DNA; 1872 BP.
XX	
XX	T28491;
XX	
XX	01-APR-1997 (first entry)
XX	
XX	P. aeruginosa detection probe #2.
XX	
XX	Detection; probe; amplification primer; bacterial pathogen; pneumonia;
XX	Escherichia coli; Klebsiella pneumoniae; Pseudomonas aeruginosa;
XX	Proteus mirabilis; Streptococcus pneumoniae; Staphylococcus aureus;
XX	Staphylococcus epidermidis; Enterococcus faecalis; respiratory tract;
XX	Staphylococcus saprophyticus; Streptococcus pyogenes; urinary tract;

KW	Haemophilus influenzae; Moraxella catarrhalis; septicaemia; meningitis;
KW	infection; intra-abdominal infection; skin infection;
KW	bacterial resistance; beta-lactam antibiotic; ds.
XX	
XX	Synthetic.
XX	
XX	
PN	WO9608582-A2.
XX	
XX	21-MAR-1996.
XX	
XX	12-SEP-1995; 95WO-CA00528.
PF	
XX	
PR	12-SEP-1994; 94US-0304732.
XX	
PA	(BERG/) BERGERON M G.
PA	(OUEL/) OUELLETTE M.
PA	(ROYP/) ROY P H.
XX	
PI	Bergeron MG, Ouellette M, Roy PH;
XX	
XX	WPI; 1996-179953/18.
DR	
XX	
PT	Method for the detection of bacterial species using probes and
PT	primers - allows detection and quantification of antibiotic
PT	resistant bacteria in patients, the environment and food
XX	
PS	Claim 34; Page 71-72; 216pp; English.

Sequence 1872 BP; 402 A; 651 C; 555 G; 264 T; 0 other;

[illegible]







CC	isolated from Thermus aquaticus.	
XX		
SQ	Sequence 1281 BP; 264 A; 425 C; 406 G; 186 T; 0 other;	
	Query Match 11.2%; Score 43; DB 21; Length 1281;	
	Best Local Similarity 46.3%; Pred. No. 0.033;	
	Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;	
QY	3 accggcaccgaagctgacatcattgggttggtggcaccggctcccgagctggagatcgcgggc 62	
DB	973 atgcgccccggggccaaatacaatacaagaacgagcgccacgctcttttgaggccaccac 1032	
QY	63 aatgcggccgacgagctgaggaagggggaagacggtccgctcgtctcgtctcgtctcc 122	
DB	1033 ggcaccgcccccaagtacgtggccaggacaaagtgaacccacgacgctcctctctcc 1092	
QY	123 tgggaactctttgatgagcagctgagatgagtaacaagagagcgtccctccctgcgcagctc 182	
DB	1093 ggggagatgatctctgctacgtggctggaacgagcgcgacccatcatcagggcc 1152	
QY	183 acagcgagatcagcatcgagcgccgggtccactctcgtggtggcagaagtacgtcggagcc 242	
DB	1153 atggagagaccatcagcaaggcgtgtcactacgacttccacgcctcctggtggcc 1212	
QY	243 caaggcaaggccattggcgcacaaagttcggcgagtgctcgtcgcgggagacatctac 302	
DB	1213 gagggcaagccgcacgcttcttaagaccagcaggttcggccaggccctgatccagcac 1272	
QY	303 aaggagt 309	
DB	1273 atggact 1279	
RESULT 12		
V35659		
ID	V35659 standard; DNA; 1480 BP.	
XX		
AC	V35659;	
DT	08-SEP-1998 (first entry)	
XX		
DE	Heat-resistant isocitrate dehydrogenase encoding DNA.	
XX		
KW	Isocitrate dehydrogenase; heat-resistant; thermotable; recombinant;	
KW	Thermus aquaticus; ss.	
OS	Thermus aquaticus.	
XX		
FH	Location/Qualifiers	
FT	180..1460	
FT	/tag= a	
FT	/product= "isocitrate dehydrogenase"	
XX		
PN	JP10165181-A.	
XX		
PD	23-JUN-1998.	
XX		
PF	09-DEC-1996; 96JP-0328400.	
XX		
PR	09-DEC-1996; 96JP-0328400.	
XX		
PA	(AGEN ) AGENCY OF IND SCI & TECHNOLOGY.	
XX		
DR	WPI; 1998-406106/35.	
DR	P-PSDB; W60606.	
XX		
PT	Heat-resistant iso:citrate dehydrogenase gene - may be prepared by	
PT	inserting it into expression vector and transforming it into host	
PT	cell	
XX		
PS	Claim 2; Pages 6-8; 9pp; Japanese.	
XX		
CC	This DNA encodes a heat-resistant (thermostable) isocitrate	
CC	dehydrogenase. A host transformed with a recombinant vector comprising	
CC	the thermostable isocitrate dehydrogenase encoding DNA can be used to	
CC	produce the protein recombinantly. The method can prepare the heat-	
CC	resistant isocitrate dehydrogenase efficiently.	
XX		
SQ	Sequence 1480 BP; 298 A; 508 C; 459 G; 215 T; 0 other;	
	Query Match 11.2%; Score 43; DB 19; Length 1480;	
	Best Local Similarity 46.3%; Pred. No. 0.034;	
	Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;	
QY	3 accggcaccgaagctgacatcattgggttggtggcaccggctcccgagctggagatcgcgggc 62	
DB	1152 atgcgccccggggccaaatacaatacaagaacgagcgccacgctcttttgaggccaccac 1211	
QY	63 aatgcggccgacgagctgaggaaggggaagacggtccgctcgtcgtctcgtctctcc 122	
DB	1212 ggcaccgcccccaagtacgtggccaggacaaagtgaacccacgacgctcctctcc 1271	
QY	123 tgggaactctttgatgagcagctcgtgatgatacaagagagcgtccctccctgcgcagctc 182	
DB	1272 ggggagatgatctctcgtacgtccctggctggaacgagggcgccacccatcatcagggcc 1331	
QY	183 acagcgagatcagcatcgagcgccgggtccactctcgtggtggcagaagtacgtcggagcc 242	
DB	1332 atggagagaccatcagcaaggcgtgtcactacgacttccacgcctcctggtggcc 1391	
QY	243 caaggcaaggccattggcgcacaaagttcggcgagtgctcgtcgcgggagacatctac 302	
DB	1392 gagggcaagccgcacgcttcttaagaccagcaggttcggccaggccctgatccagcac 1451	
QY	303 aaggagt 309	
DB	1452 atggact 1458	
RESULT 13		
V46834		
ID	V46834 standard; DNA; 2872 BP.	
XX		
AC	V46834;	
DT	10-APR-2000 (first entry)	
XX		
DE	Heat-resistant isocitrate dehydrogenase encoding genomic DNA.	
XX		
KW	Heat-resistant; isocitrate dehydrogenase; ICD; thermostable;	
KW	medical chemistry; agricultural; pharmacology; ds.	
OS	Thermus aquaticus.	
XX		
PN	JP11346780-A.	
XX		
PD	21-DEC-1999.	
XX		
PF	09-JUN-1998; 98JP-0176629.	
XX		
PR	09-JUN-1998; 98JP-0176629.	
XX		
PA	(AGEN ) AGENCY OF IND SCI & TECHNOLOGY.	
XX		
DR	(MIYA/) MIYAZAKI K.	
DR	(ORIY ) ORIENTAL YEAST CO LTD.	
XX		
DR	WPI; 2000-109695/10.	
DR	P-PSDB; Y56857.	
XX		
PT	DNA of a structural gene coding a heat-resistant isocitrate	
PT	dehydrogenase - useful as a reagent for research in medical chemistry,	
PT	biochemistry, agricultural chemistry and pharmacology	
XX		
PS	Claim 1; Page 9-12; 22pp; Japanese.	

XX The invention provides a DNA of a structural gene coding a heat-resistant  
CC isocitrate dehydrogenase (ICD) starting from initiation codon and ending  
CC at terminating codon of the 2872 base sequence (Z46834), where the second  
CC codon GCC from the Initiation codon ATG is replaced by GCT, GCA or GCG,  
CC and/or the ninth codon CCC is replaced by CCG, CCA or CCG. The ICD is  
CC useful as a reagent for researches in medical chemistry, biochemistry,  
CC agricultural chemistry and pharmacology. The method can prepare a heat-  
CC resistant ICD in large amount at low cost. The present sequence  
CC represents the genomic DNA encoding ICD.  
XX Sequence 2872 BP; 519 A; 988 C; 966 G; 399 T; 0 other;

Query Match 11.2%; Score 43; DB 21; Length 2872;  
Best Local Similarity 46.3%; Pred. No. 0.039;  
Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 3 accgcccacacgctgacatcattgggtggccaccggctccgagctggagatcgcgggc 62  
DB 1152 atcgccccggggcccaactcaactacaagacggcccgctctttgagggcaccac 1211  
QY 63 aatgcccgcagagctgaggaagaggggaagacggtcccgctgctgtctgtctctcc 122  
DB 1212 ggcacgcgcccccaagtacgctggccaggacaaggtgaaccccgagcgctcatctcc 1271  
QY 123 tgggaactctttgagcagctgagtgagtgagtgagtgagtgagtgagtgagtgagtc 182  
DB 1272 gggggagatgatctcgtctacctggctggcagggagggcgagctcatcctctcc 1331  
QY 183 acagcgagatcagcagctgagggcggtccactctcgctggcagagtgagtgagtc 242  
DB 1332 atggagaggacatcagcaaggcgtgttcacctacgactccacgctcctgtgtggcc 1391  
QY 243 caagcagagccatggtcagatcacaggttcggcgagtgctcctgcgggacgactac 302  
DB 1392 gaggcagcccgccacgcttcttaagaccagcgagttcgccaggccctgacagcac 1451  
QY 303 aaggagt 309  
DB 1452 atggact 1458

RESULT 14 ...  
ID 257437  
XX 257437 standard; DNA; 2872 BP.  
AC 257437;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE Heat resistant isocitrate dehydrogenase encoding DNA SEQ ID NO:2.  
XX  
KW Heat resistant; isocitrate dehydrogenase; ICD; ammonia; determination;  
KW ds.  
XX  
OS Thermus aquaticus.  
XX  
FH Key Location/Qualifiers  
FT CDS 180..1460  
FT /\*tag= a  
FT /product= "heat resistant isocitrate dehydrogenase"  
XX  
PN JP11346781-A.  
XX  
PD 21-DEC-1999.  
XX  
PF 09-JUN-1998; 98JP-0176643.  
XX  
PR 09-JUN-1998; 98JP-0176643.  
XX  
XX (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
PA (MIYA/) MIYAZAKI K.

PA (ORIY ) ORIENTAL YEAST CO LTD.  
XX  
DR WPI; 2000-109696/10.  
DR P-PSDB; Y67969.  
XX  
PT A liquid reagent for the elimination of ammonia - useful for the  
PT determination of biosubstances in a sample  
XX  
PS Claim 5; Page 15-18; 28pp; Japanese.  
XX  
CC The present invention describes a solution type reagent for the  
CC elimination of ammonia containing a heat-resistant isocitrate  
CC dehydrogenase (ICD) showing a high stability at an alkaline pH and  
CC responding to both NADP+ and NAD-. The reagent is used for the  
CC determination of biosubstances in a sample. The reagent can eliminate  
CC ammonia in a very short period. The present sequence encodes ICD  
CC isolated from Thermus aquaticus.  
XX  
SQ Sequence 2872 BP; 519 A; 988 C; 966 G; 399 T; 0 other;

Query Match 11.2%; Score 43; DB 21; Length 2872;  
Best Local Similarity 46.3%; Pred. No. 0.039;  
Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 3 accgcccacacgctgacatcattgggtggccaccggctccgagctggagatcgcgggc 62  
DB 1152 atcgccccggggcccaactcaactacaagacggcgcccgctctttgagggcaccac 1211  
QY 63 aatgcccgcagagctgaggaagaggggaagacggtcccgctgctgtctgtctctcc 122  
DB 1212 ggcacgcgcccccaagtacgctggccaggacaaggtgaaccccgagcgctcatctcc 1271  
QY 123 tgggaactctttgagcagctgagtgagtgagtgagtgagtgagtgagtgagtc 182  
DB 1272 gggggagatgatctcgtctacctggctggcagggagggcgagctcatcctgtgtggcc 1331  
QY 183 acagcgagatcagcagctgagggcggtccactctcgctggcagagtgagtgagtc 242  
DB 1332 atggagaggacatcagcaaggcgtgttcacctacgactccacgctcctgtgtggcc 1391  
QY 243 caagcagagccatggtcagatcacaggttcggcgagtgctcctgcgggacgactac 302  
DB 1392 gaggcagcccgccacgcttcttaagaccagcgagttcgccaggccctgacagcac 1451  
QY 303 aaggagt 309  
DB 1452 atggact 1458

RESULT 15  
ID 112068  
XX 112068 standard; DNA; 4195 BP.  
AC 112068;  
XX  
DT 08-APR-1996 (first entry)  
XX  
DE FLT4 receptor tyrosine kinase DNA.  
XX  
KW FLT4; receptor tyrosine kinase; probe; metastasis; lymphoma;  
KW lymphangioma; immunological disease; cancer; diagnosis; therapy; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 20..3976  
FT /\*tag= a  
XX  
PN W09533772-A1.  
XX  
PD 14-DEC-1995.  
XX

```
PF 09-JUN-1995; 95WO-FI00337.
XX
PR 09-JUN-1994; 94US-0257754.
XX
PA (ALIT/) ALITALO K.
PA (KAIP/) KAIPAINEN A.
PA (KARN/) KARNANI P.
PA (KORH/) KORHONEN J.
PA (MATI/) MATIKAINEN M.
PA (MUST/) MUSTONEN T.
PA (PAJU/) PAJUSOLA K.
XX
PI Alitalo K, Kaipainen A, Karnani P, Korhonen J, Matikainen M;
PI Mustonen T, Pajusola K;
XX
DR WPI; 1996-040189/04.
DR P-PSDB; R90528.
XX
XX Antibodies against FLT4 receptor tyrosine kinase, hybridomas and
PT nucleic acid probes - used to diagnose and treat e.g. metastatic
PT cancers, involving alterations to lymphatic vessels.
PT
XX Claim 9; Page 27-33; 54pp; English.
XX
CC A DNA sequence (T12068) coding for FLT4 receptor tyrosine kinase
CC (R90528) is used as a probe that specifically binds/hybridizes to
CC DNA encoding human FLT4. Such probes are used to detect FLT4, a
CC novel marker for lymphatic vessels and some high endothelial
CC venules, in biological tissue for use in diagnosis and therapy of
CC e.g. inflammatory, infectious and immunological diseases, metastatic
CC lymph nodes and lymphangiomas.
XX
SQ Sequence 4195 BP; 889 A; 1278 C; 1305 G; 723 T; 0 other;
```

Query Match 10.7%; Score 41.2; DB 17; Length 4195;  
Best Local Similarity 46.3%; Pred. NO. 0.12;  
Matches 136; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
Qy 51 gagatcgcgggcaatgcggccgacgactgtgaggaaggaggagagcggtccgcgtcgc 110
Db 2075 gacctctgtgacgtgagcgactcgtgagatcagtgcttggccggagcgac 2134
Qy 111 tcgtctctcctcgggaactctttgatgacgactcgagtgatcacaggagagcgctcctc 170
Db 2135 gcgccagcatcgtgtgtacaagacgagagcgctcgtgaggaaaagtctgagtcgac 2194
Qy 171 cctgccacgtcacagcagagatcagcatcgaggccgggtccactctcggctggcagaag 230
Db 2195 ttggcgactccaaccagaagctgagcatccagcgcgtgcgagaggatgcgggacgc 2254
Qy 231 taagtcgagcccaaggcaagccattggcatcgacaagttcggcgcgagtgctcctgcc 290
Db 2255 tatctgtcagcgtgtgtcaacgccaaagggtcgtcgtaactctccgccagcgtggccgtg 2314
Qy 291 gggacgatctacaaggaggtacgcgcatcacctggtgagagatcattgcaactgcc 344
Db 2315 gaaggtccgaggataaggcagcagatggagatcgtgatcctctgtcgtaccggc 2368
```

Search completed: November 4, 2000, 13:45:37  
Job time: 16464 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:15 ; Search time 4352.3 Seconds  
(without alignments)  
546.925 Million cell updates/sec

Title: US-09-300-482-356

Perfect score: 385

Sequence: 1 caaccggaccacgaacgtgac.....acggctctggaggttttttt 385

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est25.\*
- 26: gb\_est26.\*
- 27: gb\_est27.\*
- 28: gb\_est28.\*
- 29: gb\_est29.\*
- 30: gb\_est30.\*
- 31: gb\_est31.\*
- 32: gb\_est32.\*
- 33: gb\_est33.\*
- 34: gb\_est34.\*
- 35: gb\_est35.\*
- 36: gb\_est36.\*
- 37: gb\_est37.\*
- 38: gb\_est38.\*
- 39: gb\_est39.\*
- 40: gb\_est40.\*
- 41: em\_estba.\*
- 42: em\_estfun.\*
- 43: em\_esthum1.\*

- 44: em\_esthum2.\*
- 45: em\_esthum3.\*
- 46: em\_esthum4.\*
- 47: em\_esthum5.\*
- 48: em\_esthum6.\*
- 49: em\_esthum7.\*
- 50: em\_esthum8.\*
- 51: em\_esthum9.\*
- 52: em\_esthum10.\*
- 53: em\_esthum11.\*
- 54: em\_esthum12.\*
- 55: em\_esthum13.\*
- 56: em\_esthum14.\*
- 57: em\_esthum15.\*
- 58: em\_esthum16.\*
- 59: em\_esthum17.\*
- 60: em\_esthum18.\*
- 61: em\_esthum19.\*
- 62: em\_esthum20.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estom.\*
- 68: em\_estov1.\*
- 69: em\_estov2.\*
- 70: em\_estpl1.\*
- 71: em\_estpl2.\*
- 72: em\_estpl3.\*
- 73: em\_estpl4.\*
- 74: em\_estpl5.\*
- 75: em\_estrol.\*
- 76: em\_estro2.\*
- 77: em\_estro3.\*
- 78: em\_estro4.\*
- 79: em\_estro5.\*
- 80: em\_estro6.\*
- 81: em\_estro7.\*
- 82: em\_estro8.\*
- 83: em\_estro9.\*
- 84: em\_estrol0.\*
- 85: em\_estrol1.\*
- 86: em\_estrol2.\*
- 87: em\_estrol3.\*
- 88: gb\_gss1.\*
- 89: gb\_gss2.\*
- 90: gb\_gss3.\*
- 91: gb\_gss4.\*
- 92: gb\_gss1.\*
- 93: gb\_gss2.\*
- 94: gb\_gss3.\*
- 95: gb\_gss4.\*
- 96: gb\_gss5.\*
- 97: gb\_gss6.\*
- 98: gb\_gss7.\*
- 99: gb\_gss8.\*
- 100: gb\_gss9.\*
- 101: gb\_gss10.\*
- 102: gb\_gss11.\*
- 103: gb\_gss12.\*
- 104: gb\_gss13.\*
- 105: gb\_gss14.\*
- 106: gb\_gss15.\*
- 107: gb\_gss16.\*
- 108: gb\_gss17.\*
- 109: gb\_gss18.\*
- 110: gb\_gss19.\*
- 111: gb\_gss20.\*
- 112: gb\_gss21.\*
- 113: gb\_gss22.\*
- 114: gb\_gss23.\*
- 115: gb\_gss24.\*
- 116: gb\_gss25.\*

117: gb\_gss18:\*  
118: gb\_gss19:\*  
119: em\_gss13:\*  
120: gb\_gss20:\*  
121: gb\_gss21:\*  
122: gb\_gss22:\*  
123: gb\_gss23:\*  
124: gb\_gss24:\*  
125: em\_gss14:\*  
126: em\_gss15:\*  
127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	343.6	89.2	541	33	BE129685 945029E06
C 2	331.4	86.1	526	20	AW216325 687049D06
C 3	317	82.3	592	20	AW202478 687037D05
C 4	273.4	71.0	463	11	AI586588 486049B09
C 5	273.4	71.0	562	13	AI855050 606074D04
C 6	273.4	71.0	657	20	AW147017 707009C12
C 7	258.4	67.1	526	24	AW746961 WSL_56_CO
C 8	249.6	64.8	620	35	BE490044 WHE0364_B
C 9	246.4	64.0	581	35	BE364694 P11_14_CO
C 10	246.4	64.0	619	22	AW400161 707052H08
C 11	238.4	61.9	540	34	BE215654 HV_CEB000
C 12	237.2	61.6	705	23	AW553303 LGL_342_G
C 13	235.2	61.1	811	35	BE420889 HWM003_F0
C 14	229.6	59.6	286	35	BE344940 946030B04
C 15	226.6	58.9	747	35	BE418437 SCL033_CO
C 16	224.2	58.2	709	15	AU096596 AU096596
C 17	221	57.4	581	35	BE422449 WHE0055_H
C 18	194	50.4	483	21	AW284966 LGL_296_E
C 19	191.4	49.7	664	12	AI649526 60305G06
C 20	189.4	49.2	476	9	AI165535 A085p38u
C 21	184	47.8	458	24	AW781561 s181g05.Y
C 22	183.8	47.7	590	36	C99650 C99650_Rice
C 23	183.4	47.6	646	24	AW775545 EST334610
C 24	183.4	47.6	591	24	AW776336 EST335401
C 25	179.8	46.7	570	13	AI822218 L0_59AT3
C 26	178.2	46.3	382	24	AW734032 sk79f08.Y
C 27	175.4	45.6	464	35	BE488527 WHE1058_A
C 28	174.4	45.3	492	18	AV391286 AV391286
C 29	174.4	45.3	531	18	AV388280 AV388280
C 30	171.6	44.6	644	21	AW310464 sf36e03.X
C 31	170.8	44.4	960	33	BE053034 GA_Pa003
C 32	170.6	44.3	573	20	AW203636 sf36e03.Y
C 33	170.2	44.2	518	22	AW508848 s141b02.Y
C 34	170	44.2	484	13	AI900158 sc01f09.Y
C 35	170	44.2	526	20	AW201472 sf03e09.Y
C 36	169.2	43.9	581	14	AI966148 sc34c04.Y
C 37	166.8	43.3	484	24	AW781142 s189f08.Y
C 38	164.4	42.7	551	20	AW156809 se31b02.Y
C 39	163.8	42.5	635	18	AV524053 AV524053
C 40	162.4	42.2	651	12	AI726103 BNLGH1502
C 41	162.2	42.1	563	18	AV523269 AV523269
C 42	162.2	42.1	579	18	AV523749 AV523749
C 43	162.2	42.1	580	18	AV523732 AV523732
C 44	162.2	42.1	583	18	AV523633 AV523633
C 45	162.2	42.1	619	18	AV522729 AV522729

ALIGNMENTS

RESULT 1  
BE129685/c

LOCUS BE129685 541 bp mRNA EST 21-JUN-2000  
DEFINITION 945029E06.X2 945 - Mixed adult tissues from Walbot lab, same as 707  
(SK) Zea mays cDNA, mRNA sequence.  
ACCESSION BE129685  
VERSION BE129685.1 GI:8577048  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
AUTHORS Walbot,V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
JOURNAL University  
COMMENT Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 945029 row: E column: 06.  
FEATURES  
Location/Qualifiers  
1..541  
/organism="Zea mays"  
/cultivar="W23"  
/db\_xref="taxon:4577"  
/clone\_lib="945 - Mixed adult tissues from Walbot lab,  
same as 707 (SK)"  
/tissue\_type="tassel, kernal, silk, husk, root, leaf"  
/dev\_stage="fully-grown"  
/lab\_host="DH10B"  
/note="Organ: tassel, kernal, silk, husk, root, leaf;  
Vector: pGAD10; Site: 1: EcoRI; cDNA library from fully  
differentiated maize tissues from an active Mutator plant.  
Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,  
root, leaf). Unidirectionally cloned. New library number  
given to library 707 for additional sequencing."

BASE COUNT 125 a 169 c 131 g 116 t  
ORIGIN

Query Match 89.2%; Score 343.6; DB 33; Length 541;  
Best Local Similarity 94.9%; Pred. No. le-82;  
Matches 355; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 12 aagcctgacatcattgggtggacaccgctccgagctgagatcgcggaatgcgccc 71  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 541 AAGCCTGACCTCATGTGTGATGGGCGCCGCTCCGAGTGGAGATCGGGCGCCG 482  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 72 gacgagctgaggaaggagggaagacggttcgctgctctctctctctctctctct 131  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 481 GACGAGCTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 422  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 132 tttgatgagcagtcgagatgagatgagatgagatgagatgagatgagatgagatgag 191  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 421 TTTGATGAGCAGTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 362  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 192 atcgatcgagc 251  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 361 ATCAGCATTTGAGGCTGGGCTCCCTCTTGGCTGGCAGAGTACGTCGGAGCCAGG 302  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 252 gccattggatcgacaaagtctcgcgagtgctctctcccggaacgacatcacaagaatc 311  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 301 GCCATTGATCGACAGTTCGGCGGAGTCTCTCTCCGCGGAGCATCTCAAGAGGATC 242  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 312 ggcacatcgctggagacatcatgcaactcccaagagcttttaagagcttaacacgctc 371  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 241 GGCATCCCGTGGAGAGCATCTTTGACACTGCCAAGAGCTTTTAAGACTTAACAACGTC 182  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 372 tgggagttttttt 385  
|| | ||||| |







Db 322 TGCAGCTGCCAGAGCTTTTAAGAGCTTAACAACGCTCTGAGTTTTTTTCT 272

## RESULT 6

LOCUS AW147017 657 bp mRNA EST 03-NOV-1999  
 DEFINITION 707009C12.y1 707 - Mixed adult tissues from Walbot lab (SK) Zea  
 mays cDNA, mRNA sequence.  
 ACCESSION AW147017  
 VERSION AW147017.1 GI:6194913  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 REFERENCE 1 (bases 1 to 657)  
 AUTHORS Walbot,V.  
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 707009 row: C column: 12.

## FEATURES

source  
 1. .657  
 /location/Qualifiers  
 /organism="Zea mays"  
 /cultivar="W23"  
 /db\_xref="taxon:4577"  
 /clone\_lib="707 - Mixed adult tissues from Walbot lab (SK)"  
 /tissue\_type="tassel, kernel, silk, husk, root, leaf"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: tassel, kernel, silk, husk, root, leaf;  
 Vector: pBAD10; Site\_1: EcoRI; cDNA library from fully  
 differentiated maize tissues from an active Mutator  
 plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,  
 husk, root, leaf). Unidirectionally cloned."  
 BASE COUNT 133 a 203 c 198 g 123 t  
 ORIGIN

Query Match 71.0%; Score 273.4; DB 20; Length 657;  
 Best Local Similarity 96.2%; Pred. No. 9.1e-64;  
 Matches 280; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 caaccggcaccagctgacatcatcattgggtggccaccgctccgagctggagatcgcg 60  
 Db 367 CGACCGCCCAACAGCTTCACTTGTGATGGCCACCGCTCCGAGCTGGAGATCGCGG 426  
 QY 61 gcaatcgccgacgactgaggaaggagggaagacggtcccgctgctctgctctgtct 120  
 Db 427 CCAAGCGCCGCGAGAGCTGAGGAAGGAGGGAAGAGCGGTCCGCTCGTCTGCTGT 486  
 QY 121 cctgggaactctttgatgagcagctcgatgagtacagagagcgtctccctcgcgag 180  
 Db 487 CCTGGGAACTCTTTGATGACAGCTCGGATGAGTACAGAGAGCGCTCCCTCGCGCG 546  
 QY 181 tcacagcgagatcagcatcgagccgggtcccaactctcggtggcgagagtcgctcgag 240  
 Db 547 TCACAGCGAGGATCAGCATCGAGCGCGGTCCACACTCTCGCTGGCAGAGTACGCCGCG 606  
 QY 241 ccaaggcgaagccattgcatcgacagatctcggtcgcgagtgctcctcgcg 291  
 Db 607 CCCAGGGCAAGGCAATGGCATCGACAAGTTCCGGCGCGAGTGCTCCTCGCG 657

## RESULT 7

LOCUS AW746961 526 bp mRNA EST 19-JUL-2000  
 DEFINITION WSL\_56\_C06\_gl\_A002 Water-stressed 1 (WS1) Sorghum bicolor CDNA,  
 mRNA sequence.  
 ACCESSION AW746961  
 VERSION AW746961.1 GI:7660699  
 KEYWORDS EST.  
 SOURCE Sorghum.  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.  
 REFERENCE 1 (bases 1 to 526)  
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 ,L.H.  
 TITLE An EST database from Sorghum: water-stressed plants  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: T7  
 High quality sequence start: 73  
 High quality sequence stop: 526  
 POLYA=Yes.

## FEATURES

source  
 1. .526  
 /location/Qualifiers  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4538"  
 /clone\_lib="Water-stressed 1 (WS1)"  
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after  
 water was withheld; Vector: Lambda Zap; Site\_1: XhoI;  
 Site\_2: EcoRI; The library was made from poly-A RNA in the  
 cloning vector lambda ZAP II. Clones to be sequenced were  
 prepared by mass excision."  
 BASE COUNT 134 a 110 c 151 g 131 t  
 ORIGIN

Query Match 67.1%; Score 258.4; DB 24; Length 526;  
 Best Local Similarity 85.0%; Pred. No. 9.9e-60;  
 Matches 289; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
 QY 46 agctggagatcgcgccaatcgccgacgagctgaggaagaggaggaagacggtccgcg 105  
 Db 1 AACTGGAGATTCGGRAAAGGCTGCCGATGAGTTGAGGAAGGAGGGAAGACAGTCCG 60  
 QY 106 tcgtctgtctgtctctctctgggaactctttgatgagcagtcggtatgatacaagagagcg 165  
 Db 61 TCCTCTCATTTGTTTCTTGGGAACCTTTTCGATGAACAGTCAGATGAATACAGGAGATG 120  
 QY 166 tcctccctcgccagctcaacagcgaggatcagcagcgccgggtccactctcggtggc 225  
 Db 121 TCCTGCTCGAGCGCTCAGCGCAAGAAATCAGCATTCGAGCCGGGTCTACTCTCGGCTGC 180  
 QY 226 agaagtacgctcgagcccaagcgaagccattggcatcgacagtcggtcgagcagtc 285  
 Db 181 AGAAGTATGTCGAGAGCCAGGCGCAGGCCATTGGCATCGACAAATTCGGTGGAGTGC 240  
 QY 286 ctccgggagcatctacaaggagtagcgcatcacccgtgagagcattcattcaactgcga 345  
 Db 241 CTCCCGGAACGATCTACAGGAGTATGGCATCCTCTGTGGAGAGCGCTCATTCGGCGGCC 300  
 QY 346 agagctttaagagcttaacaacacgctgtggaggtttttttt 385  
 Db 301 AGAGCTTTTAAGAGTTAACACAACACAGTCCTTGAGTTTTT 340

QY	181	tcacagcagagatcagcatcagcagcgggtccactctcgctgcgacagagtcacgag	240
Db	239	TCACCGCAAGGATCAGCATTGAGCGCGGGTCTACCTCGGATGCAAAAGTACGTCGGGT	298
QY	241	ccacagcagagcattgctgcacacagttcgcgcgagtgctcctcgcgggacgatc	300
Db	299	CCAAGGGCAAGACCATCGGCATCGACAGTTCCGGCGCCAGCGACCTGCCCGGAAGATCT	358
QY	301	acaagagtcagcagcattcaccctggagagcatcattgcaactgccaagagctttaagagc	360
Db	359	ACAAGGAGTACGGCATCACCGCGGAGAACGTCATCGCGCAGCAAGAGCCGTGAAGATC	418
RESULT	9		
LOCUS	BE364694		
DEFINITION	P11_14_C06.g1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA, mRNA sequence.		20-JUL-2000
ACCESSION	BE364694		
VERSION	BE364694.1	GI:9306149	
KEYWORDS	EST.		
SOURCE	Sorghum bicolor		
ORGANISM	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.		
REFERENCE	1 (bases 1 to 581)		
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt,L.H.		
TITLE	An EST database from Sorghum: pathogen-induced plants		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.		
FEATURES	Location/Qualifiers		
source	1..581		
	/organism="Sorghum bicolor"		
	/db_xref="taxon:4558"		
	/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2: EcoRI; Two-week-old sorghum plants (isolate FRM421 of cultivar) were infected with pathogen (isolate FRM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."		
BASE COUNT	144 a 119 c 163 g 155 t		
ORIGIN			
Query Match	64.0%;	Score 246.4;	DB 35; Length 581;
Best Local Similarity	85.6%;	Pred. No. 1.7e-36;	
Matches 274;	Conservative	0; Mismatches 46;	Indels 0; Gaps 0;
QY	1	caaccgacacagcctgacatcattgggtggccacggcgtcccgagtcggagtcg	60
Db	59	CGACCGACACAGCCTGACTACATATAATGAGCACTGGTTCTGAGGTAGAGATTGCGG	118
QY	61	gcattgcgcccagcagctgaggaagaggaggagcgtccgctcgtcttcgctct	120
Db	119	TGAAGGCTGCCGAAGAGTGTACGAAGAGGGGAAGACCCTCCGTGTGTGTCATTGTT	178
QY	121	cctgggaactcttctgacgacgtcagtcagtcagtcagtcagtcagtcagtcagtc	180
Db	179	GCTGGGNACTCTTCGACGACGAGTCAGATGAGTACGAAGGAGGAGCGTCTCCCTGAGCGC	238
QY	181	tcacagcagagatcagcatcagcagcgggtccactctcgctgcgacagagtcacgag	240
Db	239	TCACCGCAAGGATCAGCATTGAGCGCGGGTCTACCTCGGATGCAAAAGTACGTCGGGT	298
QY	241	ccacagcagagcattgctgcacacagttcgcgcgagtgctcctcgcgggacgatc	300
Db	299	CCAAGGGCAAGACCATCGGCATCGACAGTTCCGGCGCCAGCGACCTGCCCGGAAGATCT	358
QY	301	acaagagtcagcagcattcaccctggagagcatcattgcaactgccaagagctttaagagc	360
Db	359	ACAAGGAGTACGGCATCACCGCGGAGAACGTCATCGCGCAGCAAGAGCCGTGAAGATC	418
RESULT	9		
LOCUS	BE364694		
DEFINITION	P11_14_C06.g1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA, mRNA sequence.		20-JUL-2000
ACCESSION	BE364694		
VERSION	BE364694.1	GI:9306149	
KEYWORDS	EST.		
SOURCE	Sorghum bicolor		
ORGANISM	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.		
REFERENCE	1 (bases 1 to 581)		
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt,L.H.		
TITLE	An EST database from Sorghum: pathogen-induced plants		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.		
FEATURES	Location/Qualifiers		
source	1..620		
	/organism="Triticum aestivum"		
	/cultivar="Chinese Spring"		
	/db_xref="taxon:4565"		
	/clone="WHE0364_B03_C06"		
	/clone_lib="Wheat cold-stressed seedling cDNA library"		
	/tissue_type="Seedling"		
	/dev_stage="Five-day old seedling"		
	/lab_host="E. coli SOLR"		
	/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site.1: EcoRI; Site.2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were transferred to 5 C cold room and kept for 48 hr. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."		
BASE COUNT	155 a 156 c 176 g 133 t		
ORIGIN			
Query Match	64.8%;	Score 249.6;	DB 35; Length 620;
Best Local Similarity	80.8%;	Pred. No. 2.4e-57;	
Matches 291;	Conservative	0; Mismatches 69;	Indels 0; Gaps 0;
QY	1	caaccgacacagcctgacatcattgggtggccacggcgtcccgagtcggagtcg	60
Db	59	CGACCGACACAGCCTGACTACATATAATGAGCACTGGTTCTGAGGTAGAGATTGCGG	118
QY	61	gcattgcgcccagcagctgaggaagaggaggagcgtccgctcgtcttcgctct	120
Db	119	TGAAGGCTGCCGAAGAGTGTACGAAGAGGGGAAGACCCTCCGTGTGTGTCATTGTT	178
QY	121	cctgggaactcttctgacgacgtcagtcagtcagtcagtcagtcagtcagtcagtc	180
Db	179	GCTGGGNACTCTTCGACGACGAGTCAGATGAGTACGAAGGAGGAGCGTCTCCCTGAGCGC	238







Sat Nov 4 18:11:32 2000

```
Qy 65 tgcggccgacgagctgaggaagagggaagacggtccgcgctcgtctcgtctcctc 124
Db 601 GNCCTGCCGAAGAGTTGACGAAGAGGGGAAAGCCCGTCCGTTGTTGTCATTGTTGCTG 542
Qy 125 ggaactctttgatgagcagtcggatgagtacaaggaagagcgtccctgcccacgtcac 184
Db 541 GGAACCTCTTCGACGAGCAGTCAGATGAG-ACAAGGAGAGCGTGTCCCTGAGGCCCGTCAC 483
Qy 185 agcggatcagcatcagcggcgggtccactctcggctggcagaagtacgtcggagccca 244
Db 482 CGCAAGGATCAGCATTCAGGCCGGGTCTACCTCGGATGGCAAAAGTACGTCGGGTCCAA 423
Qy 245 agcaaggccattggcatcgacaagtctcgcgcgagtgctcctgcccgggacgatctacaa 304
Db 422 GGGCAAGACCATCGGCATCGACAAGTTTCGGCGCCAGCGCACCTGCCCCGGAAGATCTACAA 363
Qy 305 ggagtacggcatcaccctggagagcatcattgcaactgcccaagagcttttaagagc 360
Db 362 GGAGTACGGCATCACCGCGGAGAACGTCTATCGCCGCGAGCCCAAGAGCCTGTAGATC 307
```

Search completed: November 4, 2000, 11:52:19  
Job time: 18256 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:36:23 ; Search time 4075.18 Seconds  
(without alignments)  
293.660 Million cell updates/sec

Title: US-09-300-482-311

Perfect score: 274

Sequence: 1 ctcatctcttctccacc.....gccaaagcatgccgtgcaccg 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pl1.\*

8: gb\_pl2.\*

9: gb\_pl1.\*

10: gb\_pr2.\*

11: gb\_pr3.\*

12: em\_fun.\*

13: em\_hum1.\*

14: em\_hum2.\*

15: em\_in.\*

16: em\_om.\*

17: em\_or.\*

18: em\_ov.\*

19: em\_pat.\*

20: em\_ph.\*

21: em\_pl.\*

22: em\_ro.\*

23: em\_sts.\*

24: em\_sy.\*

25: em\_un.\*

26: em\_vi.\*

27: gb\_htg1.\*

28: gb\_htg2.\*

29: gb\_in1.\*

30: gb\_in2.\*

31: em\_ba1.\*

32: em\_ba2.\*

33: em\_hum3.\*

34: em\_hum4.\*

35: gb\_pr4.\*

36: gb\_htg3.\*

37: gb\_htg4.\*

38: gb\_htg5.\*

39: gb\_htg6.\*

40: gb\_htg7.\*

41: em\_htg1.\*

42: em\_htg2.\*

43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_vil.\*  
94: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	% Match	Query Length	DB ID	Description
c	1	81.4	29.7	6653	8	ATAC016972
	2	81.4	29.7	8497	7	AC006200
	3	68	24.8	1118	5	A93589
	4	62	22.6	1084	7	AF154658
c	5	54	19.7	93234	8	ATAC009465
	6	54	19.7	95310	8	ATAC011437
c	7	47.4	17.3	37898	2	SCE34
	8	46.6	17.0	36224	2	SCD78
	9	46.4	16.9	1288	5	AR007563
c	10	46.4	16.9	161266	8	AP001278
	11	45	16.4	10834	1	AE004020
	12	44.8	16.4	3875	2	RRPHAS
						AC016972 Arabidops
						AC006200 Arabidops
						A93589 Sequence 1
						AF154658 Nicotiana
						AC009465 Arabidops
						AC011437 Arabidops
						AL353862 Streptomy
						AR007563 Sequence
						AP001278 Oryza sat
						AE004020 Xylella f
						X66407 R.ruber ORF

```

13 44.6 16.3 40790 2 MTCY493
14 44.6 16.3 77534 1 AF235504
15 44.6 16.3 182756 7 AC007789
16 44.4 16.2 43632 2 SC6A5
17 44.2 16.1 1593 1 AF113605
18 44 16.1 2647 45 OSHPRGP
19 44 16.1 28732 1 AF082100
20 43.8 16.0 273 2 MTU013938
21 43.6 15.9 154561 8 AP002819
22 43.2 15.8 1230 45 AJ242540
23 43.2 15.8 19830 2 SC3F9
24 43 15.7 701 5 A64741
25 43 15.7 38734 2 SC10A5
26 43 15.7 68848 2 MTV043
27 42.8 15.6 32274 2 SCC8A
28 42.8 15.6 133185 49 AC023240
29 42.6 15.5 7836 1 AF047717
30 42.4 15.5 989 7 AB027423
31 42.4 15.5 996 7 AB027422
32 42.4 15.5 1167 8 D55714
33 42.4 15.5 1416 7 AB027421
34 42.4 15.5 14472 1 AE003928
35 42.4 15.5 18959 1 AE004032
36 42.4 15.5 20956 1 AE004082
37 42.2 15.4 324 1 AF046836
38 42.2 15.4 1187 5 AF046840
39 42.2 15.4 1187 5 AF046840
40 42 15.3 771 5 A96762
41 42 15.3 820 53 AC059012
42 42 15.3 2157 2 PSEPTA
43 42 15.3 9344 2 D86947
44 41.8 15.3 833 5 A37833
45 41.8 15.3 833 5 AR069871

```

Z95844 Mycobacteri  
AF235504 Streptomy  
AC007789 Oryza sat  
AL049485 Streptomy  
AF113605 Streptomy  
X61280 O.sativa ge  
AF082100 Streptomy  
U13938 Mycobacteri  
AP002819 Oryza sat  
AJ242540 Volvox ca  
AL023862 Streptomy  
A64741 Sequence 1  
AL021529 Streptomy  
AL022004 Mycobacte  
AL356892 Streptomy  
AC023240 Oryza sat  
AF047717 Streptomy  
AB027423 Oryza sat  
AB027422 Oryza sat  
D55714 Oryza sativ  
AB027421 Oryza sat  
AE003928 Xylella f  
AE004032 Xylella f  
AE004082 Xylella f  
AF046836 Marine pr  
AF046840 Marine pr  
AR007559 Sequence  
A96762 Sequence 79  
AC059012 Giardia i  
D50642 Pseudomonas  
D86947 Pseudomonas  
A37833 Sequence 3  
AR069871 Sequence

## ALIGNMENTS

```

RESULT 1
ATAC016972 66653 bp DNA PLN 18-JAN-2000
LOCUS Arabidopsis thaliana chromosome I BAC F23N20 genomic sequence,
DEFINITION complete sequence.
AC016972
AC016972.5 GI:6714311
HTG.
thale cress.
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 66653)
Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B., Wu,D.,
Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E.,
Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome I BAC F23N20 genomic sequence
Unpublished
2 (bases 1 to 66653)
Lin,X. and Kaul,S.
Direct Submission
TITLE Submitted (09-DEC-1999) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 66653)
Lin,X.
Direct Submission
TITLE Submitted (18-JAN-2000) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr., Rockville, MD 20850, USA
On Jan 18, 2000 this sequence version replaced gi:6630537.
COMMENT Address all correspondence to:at@tigr.org

```

BAC clone F23N20 is from Arabidopsis chromosome I and is near the molecular marker m453. The orientation of the sequence is from SP6 to T7 end of the BAC

clone.  
Genes were identified by a combination of three methods: Gene prediction programs including GAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netpene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

## FEATURES

```

source
1..66653
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="I"
/map="m453"
/clone="F23N20"

misc_feature
1..17670
/note="overlap with BAC clone F15H11
(AC008148:92846..110514)."

misc_feature
1..15580
/note="the annotation for genes within this region can be
found in the overlapping bac F15H11 sequence 92846-108425"

misc_feature
1
/note="37813 nt before this point were not included in the
submitted sequence, due to overlap with another BAC
(F3117)"

repeat_region
complement(28..118)
/rpt_family="TAAAAA)n"

repeat_region
complement(1088..1149)
/rpt_family="POLY_A"

repeat_region
complement(3778..3842)
/rpt_family="(GAAAA)n"

repeat_region
complement(3778..3821)
/rpt_family="POLY_A"

repeat_region
complement(4015..5903)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."

repeat_region
complement(4183..4303)
/note="ATREP3|ATREP3 a consensus."

repeat_region
complement(4242..4477)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."

repeat_region
complement(4245..4396)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."

repeat_region
complement(4404..4475)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."

repeat_region
complement(4461..4600)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."

repeat_region
complement(4526..4597)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."

repeat_region
complement(4613..4670)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."

repeat_region
complement(4662..4832)
/note="ATREP4|ATREP4 An AT-rich repetitive sequence - a
consensus, from T23015 64288 to 65702 1415 nt, also
present in intron 8 of Arabidopsis thaliana gene for

```





overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tldb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/GENSCAN.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are annotated by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F623, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: [at@tigr.org](mailto:at@tigr.org).

FEATURES	source	Location/Qualifiers
misc_feature		1..84957
		/organism="Arabidopsis thaliana"
		/cultivar="Columbia"
		/db_xref="taxon:3702"
		/chromosome="II"
mRNA		<1..2020
		/note="Sequence from clone F23H14"
		complement(join(<463..924,1053..1163,1265..1417,1513..1572,1662..1967,2287..2435,2515..2584,2718..2921,3271..3465,3673..3945,4026..4187,4325..4455,4541..4700,4817..5005,5145..>5219))
		/gene="At2g01130"
		complement(<463..>5219)
gene		/gene="At2g01130"
		/codon_start=1
		/note="F10A8.1"
		complement(join(<463..924,1053..1163,1265..1417,1513..1572,1662..1967,2287..2435,2515..2584,2718..2921,3271..3465,3673..3945,4026..4187,4325..4455,4541..4700,4817..5005,5145..>5219))
		/gene="At2g01130"
CDS		/codon_start=1
		/product="putative ATP-dependent RNA helicase A"
		/protein_id="AADI4515.2"
		/db_xref="GI:6598524"
		/translation="MLDSRTSLPAFKQKHSVLTALISQNVIVISGTCGKQTQIPQF

misc_feature		2021..>84957
		/note="Sequence from clone F10A8"
		complement(join(6441..7261,7365..7474,7570..7839,7952..>8107))
		/gene="At2g01140"
		complement(6441..>8107)
gene		/gene="At2g01140"
		/note="F10A8.2"
		6543..6582
		/rpt_family="CAAAA)n"
		complement(join(6622..7261,7365..7474,7570..7839,7952..>8107))
repeat_region		/gene="At2g01140"
		/codon_start=1
		/product="putative fructose-bisphosphate aldolase, plastidic form"
		/protein_id="AADI4543.1"
		/db_xref="GI:4262250"
CDS		/translation="MASASFVKPNTLSPWIKQSFQFAHTSSASSPPRVSFAIRAGAY
		SDELVTAKSIASPGGILAIIDESNATCGKRLAIGLDNEDNRQATRQLLLITPGLG
		DYISGILFEETLYQSTKDTFVDRDANIVFGIKVDKGLSLAGSNEESWCQGLD
		GLASAEYKQGARFAKWTVPVPGPSALAVKEAAGLARYAAISQDGLVPIVE
		PEILLDGHPIERTLEAVKVMSEVFFLAQNVMFEGILLKPSMTVPQDNKRNKASP

repeat\_region

mRNA

gene

CDS

repeat\_region

tRNA

gene

mRNA

gene

CDS

repeat\_region

mRNA

gene

CDS



[illegible]





Db 91794 CTTCTCCTCGACCTCGCTTCACTCTCGTCAAGCTCAATCCGTCGCTTTCACAGG 91853  
 QY 172 acgatttgagaaatccgcctacaaagccgtcgatgagtcagtgagtcgcgcgcatggtcc 231  
 Db 91854 ACGATTGGAAGCTCGCGGGGAGAAAGCTGTGGAGGCAATTAAACCTGGATGGTTC 91913  
 QY 232 tcggcttaggcacgcgtccaccgccaagcgcgcgtcgacc 273  
 Db 91914 TAGGGCTCGGAACCGGATCCACCGCAGCTTTCGCTGTGTGATC 91955

## RESULT 7

SCE34/c 37898 bp DNA BCT 01-JUN-2000  
 LOCUS Streptomyces coelicolor cosmid E34.  
 DEFINITION AL353862  
 ACCESSION AL353862.1 GI:7649483  
 VERSION  
 KEYWORDS adenosylhomocysteine; integral membrane protein; L-lactate permease; mannose-6-phosphate isomerase; nucleotide phosphorylase; oxidoreductase; phosphomannomutase; rmbd; secreted protein; transcriptional regulator; transport protein; whiB.  
 Streptomyces coelicolor A3(2).  
 SOURCE Streptomyces coelicolor A3(2)  
 ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces. 1 (bases 1 to 37898)  
 AUTHORS Redenbach, M., Kieser, H. M., Denapaita, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D. A.  
 TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)  
 MEDLINE 97000351  
 REFERENCE 2 (bases 1 to 37898)  
 AUTHORS Saunders, D. C. and Harris, D.  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 37898)  
 AUTHORS Cerdeno, A. M., Parkhill, J., Barrell, B. G. and Rajandream, M. A.  
 JOURNAL Direct Submission  
 TITLE Submitted (25-APR-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
 COMMENT Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)  
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.  
 CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence

overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
 Cosmid E34 Overlaps cosmid E33 on the AseI-E genomic restriction fragment.

## FEATURES

source Location/Qualifiers  
 1..37898  
 /organism="Streptomyces coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid E34"  
 misc\_feature 1..161  
 /note="nominal overlap with Streptomyces coelicolor cosmid STE33"  
 complement(<1..105)  
 /note="probable 23S rRNA (fragment), coordinates taken from similarity to SW:SARRND (EMBL:M27245) Streptomyces ambifaciens rrmD gene cluster encoding 16S, 23S and 5S ribosomal RNAs"  
 complement(<1..2116)  
 /note="Coordinates taken from similarity to SW:SARRND (EMBL:M27245) Streptomyces ambifaciens rrmD gene cluster encoding 16S, 23S and 5S ribosomal RNAs"  
 /label=rrmD  
 complement(359..2116)  
 /note="probable 16S ribosomal RNA. Coordinates taken from similarity to SW:SARRND (EMBL:M27245) Streptomyces ambifaciens rrmD gene cluster encoding 16S, 23S and 5S ribosomal RNAs"  
 complement(2470..3477)  
 /gene="SCE34.01c"  
 complement(2470..3477)  
 /gene="SCE34.01c"  
 /note="SCE34.01c, possible integral membrane protein, len: 335 aa; similar to TR:O69662 (EMBL:AL022121) Mycobacterium tuberculosis hypothetical 35.5 kD protein MTU025.042c, 330 aa; fasta scores: opt: 1119 z-score: 1285.3 E(-): 0; 51.8% identity in 334 aa overlap. Contains possible hydrophobic membrane spanning regions"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative integral membrane protein"  
 /protein\_id="CAB88904.1"  
 /db\_xref="GI:7649484"  
 /translation="MGLDVFVSAHRAEMDRDLALLRRRLRTGTPETDELVALYQRTAT HSLIQSADPDQLTGRLSQVARARSVVTGTRASRDVTRFLTQCPAAYVRAHWA WPTALLSTAVALLGWTHGHPQSTIAAPSELRELTRPGGOYETVYVSNPAAFA AQVNTNAAALCLILGVFLPVLIFLQNLNLGVFGLSAGRLDTFLGLVLP HGLLELTAVFAAGTGLRLWTLIDPGFTRTRIALAEGRAAIGMVGALVLVFGA IEGVFTVSGLPTWARITIGVLAELAFIAYVVLGGRAVRDGTGDVEAERSATVPTA A"  
 complement(3489..3494)  
 RBS 3589..3595  
 RBS 3604..4560  
 gene /gene="SCE34.02"  
 3604..4560  
 CDS /gene="SCE34.02"  
 /note="SCE34.02, possible integral membrane protein, len: 318 aa; similar to TR:O69663 (EMBL:AL022121) Mycobacterium tuberculosis hypothetical 33.2 kD protein MTU025.043, 310 aa; fasta scores: opt: 734 z-score: 602.4 E(-): 4.3e-26; 42.9% identity in 319 aa overlap. Contains possible hydrophobic membrane spanning regions"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative integral membrane protein"  
 /protein\_id="CAB88905.1"  
 /db\_xref="GI:7649485"  
 /translation="MSELVTGEAVALELRPARLPSRALAVLLDLAVAVYVAVTIAL MVATASLDVAQAQTATLVLVGGPIAVETLSHRSGLKMACGLRVVRDGGPFR FRISLVRLGIVIEILMTLQVAVSLAVSRLRGRLGDVFAGLTVVRRVRFPSAGFM PPPPWLAGRFGSLDLSAVPDDLAVRQLARMGOLDPRVGMARERLAADAAATG APVPREVPFPAYLAALVQERAEARFGASAGTAAWAVPAPAPAPAPVPTAP PPVPTAPPTAPPFPSPGGLEVPDRPDRPPGGVFPVPPA"





initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
**IMPORTANT:** This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid D78 lies to the right of Q11 on the AseI-D genomic restriction fragment.  
 Location/Qualifiers

## FEATURES

source

1. 36224

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid D78"

1. 817

/gene="SCD78.01"

&lt;1. 817

/gene="SCD78.01"

/note="SCD78.01, ftsK /spoIIIE family protein, partial CDS within putative integrated plasmid, len: 271 aa; similar to many members of the ftsK /spoIIIE family e.g. SPI\_STRAZ (EMBL:558719) sporulation-regulatory protein from Streptomyces azureus plasmid pSA1.1 (303 aa), fasta scores: opt: 736 z-score: 1014.5 E(): 0, 49.0% identity in 247 aa overlap and TRSA\_STRAM (EMBL:Z19593) transfer protein trasa from Streptomyces ambofaciens plasmid pSAM2 (306 aa), fasta scores: opt: 222 z-score: 676.5 E(): 2e-30, 47.8% identity in 203 aa overlap. Also similar to SC6A9.34, a protein within another putative integrated plasmid (on Streptomyces coelicolor cosmid St6A9)(E): 5.4e-24, 45.7% identity in 247 aa overlap). Contains PS00017 ATP /GTP-binding site motif A (p-loop).  
 /codon\_start=2

/transl\_table=11

/product="ftsK /spoIIIE family protein"

/protein\_id="CAA22206.1"

/db\_xref="GI:3928710"

/translation="IGALETGGAWNNLRVLPWHLIAGATRSCKSTLLARVITQAPQ PVALVGTDCRGMEGLFADRLSALATRSREAVALTALVVDIQRMSACRTAGVPSV WELPDKRPVGVVVLVDIEALVLSDTQSKSEAEQCSLTLLRLAQALGALHLVW AGORVSDLPGLVTAARQALGGRIHCHRVDPGTAEMTGLDNKDAVAQAITAQERG VAVCTGPDGWARASHLPTDEAVATARKHSMTPELPALDRAUVALEGGCK"  
 1. 4443  
 /note="possible integrated element"

/codon\_start=2

/transl\_table=11

/product="ftsK /spoIIIE family protein"

/protein\_id="CAA22206.1"

/db\_xref="GI:3928710"

/gene="SCD78.01"

/gene="SCD78.02"

/gene="SCD78.02"

/note="SCD78.02, small hydrophobic protein, len: 65 aa; within putative integrated plasmid"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein SCD78.02"

/protein\_id="CAA22207.1"

/db\_xref="GI:3928711"

/translation="MISEGTATGFAVVGIIITVLLVRSRDVRGWEAVCVGLFGLVIGQ TPVLETFHGLVTVWISGFSHT"

1308. 2666

/gene="SCD78.03"

1308. 2666

/partial

/gene="SCD78.03"

/note="SCD78.03, possible plasmid replication initiator protein within putative integrated plasmid, len: 452 aa; similar to REPS\_STRAM (EMBL:Z19594) replication initiator protein from Streptomyces ambofaciens plasmid pSAM2 (459 aa), fasta scores: opt: 712 z-score: 983.7 E(): 0, (43.7% identity in 467 aa overlap). Note that there is no possible initiation codon upstream of the start of homology (or close downstream). The CDS given here has an arbitrary start"

misc\_feature

misc\_feature

gene

CDS

gene

CDS

gene

CDS

/codon\_start=1  
 /transl\_table=11  
 /product="putative plasmid replication initiator protein"  
 /protein\_id="CAA22208.1"  
 /db\_xref="GI:3928712"  
 /translation="MRHLPEADRDALRIADPDFRWWLOQITATGCAHPVHLSGHTT TLDNATGEILRHVDTRNEPGERLLVRCNRRSTVCGPCSLRHAGDTFLVLRAGLLGGK NVPASVTRPRLFTVLTAPSGAVHWRAGOTCPRRDGGACGHRPLGCGCTVHAPDAPA VGQPLCPDXYDTAHVLAHASKLWDREVIDVRRLASSAGIVSRFAHKLHLSFAR IAEYOKRAAVHVTVVYRLDGGPDGPDPEPAWGTDRDLRAVHASQRVTRTPYSLAT GELELSWGQTQDVCVYLHAGLGDGDDDAVAAVYAKYVSKGTDEGTGATDHKVTTCDDID SARVSRHVTRLMTHTCRLGLPEYALRLRTWTHTLGRGHILTKSRAYSTTYAALRA QRAHHGHGTTDTDAITDAHWRYVSGHTFGAALIAAGVAEDIAQNAARVDAALSAGGG AT"  
 complement(2805..4292)  
 /gene="SCD78.04c"  
 complement(2805..4292)  
 /note="SCD78.04c"  
 /gene="SCD78.04c, possible integrase/recombinase within putative integrated plasmid, len: 495 aa; similar to TR:Q37839 (EMBL:D38173) ORF469 protein (possible site-specific recombinase) from Actinophaga R4 (469 aa), fasta scores: opt: 261 z-score: 320.1 E(): 1.4e-10, 29.1% identity in 306 aa overlap and TR:O06604 (EMBL:Z95586) MTCY336.18 (RV1586c) possible integrase from M. tuberculosis integrated phage phiRv1 (469 aa), fasta scores: opt: 371 z-score: 270.8 E(): 7.8e-08, (26.6% identity in 489 aa overlap). Probable coiled-coil from 414 to 451 (38 residues) Max score: 1.608 (probability 0.99)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative integrase /recombinase"  
 /protein\_id="CAA22209.1"  
 /db\_xref="GI:3928713"  
 /translation="MLAADPNLKCVCVYARISFGRVKDAHGIEDQHRDMSAARRFG WLLVRYTNDKSKASKESVSRDDFPELLADLAAGTPEGVPVHGVMVNDRLVRRPS DMERLYKTSODGRVYHDSNGLDLYAEFGFGLGVAMSLSETRKKRRSRNSHR SRAIRGQSVAMRPFGEWDEKVTLPDPAEAIATVADHVIAGASITETRWKAGFI TSRGNFQYGTVPKVLVNRKLLGTVREIKGEIVRGDDQPIVGWEAIVTPKQWFAVTA KIREHGHTGTGPGGLVHVKLLTNILRCGNVLEDTCVNNKMGITKANVDPKLYOHAYM CKTKVDGCKNTYKRGDKTKTIEELVIAKLERDAATKAQDVPDWDKEALERALQSR RELRRWHDDETDIDDEAFRNLPVLEERIKELRVOKAHEALKEAEAEADIRKS WGAKTLTKREAMKVLGAVIALPGGKNGKTFDPLLPKPVWKTSE"  
 4444. 4575  
 /gene="SCD78.05"  
 /note="thic"  
 4444. 4575  
 /partial  
 /gene="SCD78.05"  
 /note="SCD78.05, thic', probable thiamin biosynthesis protein thic, partial CDS, len: 43 aa; similar to the extreme C-terminus of e.g. THIC\_ECOLI thiamin biosynthesis protein thic (631 aa), fasta scores: opt: 97 z-score: 137.7 E(): 2, 46.2% identity in 39 aa overlap and THIC\_MYCTU (547 aa), fasta scores: opt: 143 z-score: 272.4 E(): 6.4e-08, 53.1% identity in 49 aa overlap. Note that this gene appears to have been interrupted by the integration of the upstream element"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative thiamin biosynthesis protein thic"  
 /protein\_id="CAA22210.1"  
 /db\_xref="GI:3928714"  
 /translation="SMKISQDIRRQHGSSKGEIEGMAQSKFEFAAAGNRVYLPIAD" 4655. 6103  
 /gene="SCD78.06"  
 4655. 6103  
 /gene="SCD78.06"  
 /note="SCD78.06, unknown, len: 482 aa; similar to 17.0%; Score 46.6; DB 2; Length 36224; Best Local Similarity 52.9%; Pred. No. 9; Matches 100; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Query Match  
 Best Local Similarity 52.9%;  
 Matches 100; Conservative 0; Mismatches 89; Indels 0; Gaps 0;



```

CDS
join(25864..25872,26022..26097,26354..26802)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA92195.1"
/db_xref="GI:7106510"
/translation="MSKKTWIKADLIGRACKSELAIVATIKIGKMSDIDAECTLTIV
VKIVDPVRIIVKLKCKFSACIVSVEDDKPEEKPKDCKEAEKLEKAWKECEKCN
VGLKPGCSCSTPCSPPIICGHDGICLPPCPPCLPPCPPCLPPCPPCPGRTG
YGVYERYPGGECVIO"
complement(join(27737..27749,27842..28764))
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA92196.1"
/db_xref="GI:7106511"
/translation="MPCGYVPILSTLDFARRGTFASHQYDTGPFYDRIIRRLSFD
FTKSQVKEKURLUKKTLCRAKMASSPHAAAGFAFRPHEGAIYDLARHWPALK
RDGASDDDDINPAAAAATAVMTPVAMEDFGGSAPTPTTPRGGRRVRRMAQE
KALPSPALSTDGAHOEPLVAAMETLPQIAPLPVSETEMPVITANGANERAV
RSVLRLKEFTIFSFAVVGOTGPGMLMGFGGAGLNADTAGLGFGLAGLPGVPGAD
RWROHQILELVYLKRTLEVQVTRALDELSVCH"
join(28965..29024,29098..30608,31416..32066,32089..32627,
32682..33058)
/notes="EST C98429 (E0138) corresponds to a region of the
predicted gene.; hypothetical protein"
/codon_start=1
/protein_id="BAA92197.1"
/db_xref="GI:7106512"
/translation="MQPDRQINTHSADSNPQTGVCATPFYCRIVHSTVGAPGRG
VELPEKDQIDTRNGEINLDSNLOGAAPTDLRPPRHGHMGAAAYSLIGEISC
MHELIQDINQIAQYAAVGRRLRHPACLRRLRLVSLPTRLAYAAAVGRRLRHRP
ASVAAEWCLRHGMPIMPLLVGYDFTGLPPSPGASAYTAGLLRCRCWASTTSP
PACLRRLRVPPTRLAYAAVGRRLRHLRLASVAAEWCLRLHGHPIMLLVGYDF
TGLPSPPIGASAYTAGLLRCRCWSTTSPACLRVRVPPPARLAYATVVGRLH
LHRPASVADWCLRLHGMPIMLLVGYFTFGLRPPSPPTGADWCLHGMPTWPCLLG
TSPPACLRRLRVPPTRLAYATVVGRLHRLHRLPASVADWCLHGMPTWPCLLG
VYITAGLPPSPPIGASAYTAGLLRCRCWSTTSPACLRVRVPPPARLAYATVVG
GLHRLHRLPASVADWCLHGMPTWPCLLG
TAGLLRCRCWSTTSPPTGLPSPPTGVSACTAGLYAAVGRRLRHPACLRRLRVFP
PAMLAPMPLLVGYFTDRLFAAADCWCFRLHGLIMLLVGYFTDRLFAAADCW
PRLHGMPIMLLVGYFTDRLFAAADCWCFRLHGMPIMLLVGYFTDRLFAAADCW
CLYAAVGRRLRHPACLRRLRVLELRLAFYAAVGRRLRHPASVADWCFRLH
GNMPLLVGYFTDRLFAAADCWCFRLYGMPIMLLVGYFTDRLFAAADCWCLR
LHGMPTWPCLLG
LTFIAATTDGHLRLPSPDIAAADCWCFRLHNCASSLPVCFCRLVTRSLQVNFV
IMVERLHLYLRHWOTLLLLLRKFCYFTNHDVFNLDISACDAVLFYFNKLSNQDF
YLDIENTYLGSSNDY"
complement(33441..33596)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA92198.1"
/db_xref="GI:7106513"
/translation="MKMKWIVDDKRNHDEDDVDPANSSVATNDNDKAHQHLHS
RSTVVKV"
join(35588..35983,36642..37132,39680..39857,40268..40411)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA92199.1"
/db_xref="GI:7106514"
/translation="MRSSSSVQSRNRSLESSAPAGCGCGGGGGGIAVGGSGAWG
AAWSGLEFSLSPSSSSSAGEEWEAAASSAAVYVVVVVYVGMGLRGLGNQ
VSDRTDQAGWRIITAFSSVGSULTALITPHVIDISYSSGGAFAEFESAGSD
HAPISSPAASSTAKTSPASCRTAGRRRQWGAEEEDGVATAGGEARRQPGG
GERRRREAGEGEFVYVAVGVVSWWKKGRDDHVPVQPLPAVGRKANA
DIVAAAAALYALTGLRDRARKRYNKTNTSVGMSDNTSVLPDLYLESYETIPVLSYL
ESYDKSRQNDKHNHREIILCSNICTDLNTRCDTSKLYTPESFFYSARTPHTRQ
NETPLDLNPN"
complement(join(47013..47907,47926..48720,48802..49137,
49409..50012,50104..51310))
/notes="Similar to Arabidopsis thaliana chromosome II BAC
T13P21 genomic sequence, putative helicase. (AC006067)"
/codon_start=1
/protein_id="BAA92200.1"
/db_xref="GI:7106515"

```

```

/translation="MLANLLRFDGRRSRRLQIRSYNSLFAFTSLGADVKSINNG
TAPCFKINGIIVHRIIGSLPQOGAPPKFAQLIYDTENEVSNRMRVPDKESDPEPD
PFTVTEGAMLDENHDLVKSFRFAKRLKDHGDEKIALRLCNSKDEVOYNLPISGE
IAGILGVDSNDKYTYDVVYVQSCDSRLRVSAHPYSMAQYLPPLFYGERGFGILGK
YDFPSITGTSRYTMBEYIRLHRLUNKPNPYTCGRLSDSICVDMYSTVGSRL
KIJADNPDLRSECQGIADAIDHGLESDSGQKYLVPSSFTGGRRYVQYQADAMA
VCRVFGSPDLFTFTCSNKQWIEYDALLFEPQGVPSDRSMIVRVNMKNVEFIIDIR
EKTFTGPVLAVEFGKRLPHICLWLAANTTDFNAATVDGESSAEIPDVFDPILGYA
LVDEFMIGPCGEKNSKSCVMKNOCSEHFKPSFODETIDMEFGFTVYVRNNGRYV
KNGIILDNRWVVPYNNKLLKYOAHINVEWCKSNMIKLYFKYTKGSDRTKYFETT
GNTANKPTNSAAPNEIDEIYNARFLSTCESIWRTELEMIYISPVAGELYLRMLLI
VKGVMCFADIRTYEGVYPTFQACAEAGLLENDNEHLLFDEAIVSASSVQLRQLFY
TVVMSCTGNVRSFLDKWYTTDDIOHRVRKMGFNIDDDYDLPQSTARTDDDSGNRM
VSEELALDSVALAAHADSITPKLNSDQKRVFTIMRVNESPKEFFYVHGHTGKTFT
LACALISKVRSEKTIIVAVSSVASLILPCRTAHSRFPIDINENSLCTIKGTM
LNEALQKTSITWDEAPWTHRRCFEALORTLDDLSEHAPSNGLVPFEGKVVYLGDF
ROILPVKGSASIIASITNSPLWSHVALKLVNMRLOCNGEQQEQLKFAQ
WYALGDGKNDESEATIDIDDLIKITGDKHISIVNEVFPFASKYTFDPSLASCA
IYVCPNNSVDDINDRMVDMVPEVKEYLSCDITKSESEHIFDILLYPTFTNLSISAN
NPPAHKLLKKGVTYMLLRNQMGLCNGTRLALLSLGQRLLECOILGTNIGDRVF
IPRIALTTPSPKPTFLORQPPVRCVAMTINKSOGOTLKRGVYLKAVFTHGOLY
VAVSSTSRDGLRILTEGDDEACSSKTRNVPVYHEVQLVDMVHVLFDLLAVTHEYL
KFQOSNTLLRFGVGC"
join(53601..53816,54051..54343,54734..54829,58868..59058,
59090..59322,60053..60211,60746..60938,61620..61810,
61844..62084,62459..62754,64552..64605,64941..65004,
65106..65260)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA92201.1"
/db_xref="GI:7106516"
/translation="MSGANPKRRSIHNPCTRIHQNSSQGERKTYGAAVHLVLMAC
ATAKRGKTGAHSTSEPNQOQRAELTNRMHVHQDFGINORIISOLLMSVLGERAPS
RNTHIGKLCGDPWCTTIVGEPRIDVSHTEROCIVLSKSGRCTKRAQCAQKKLPST
GAPTKVTRYVMSFNRLGILIRVHOVSDSICLONCNSCLLSCSSCDVNPNEI
DAAGLSAPTAAPVSTGAPSSILLSPALPCPCREEEKETGAARTHRGPPA
SNAPYRRPDEADVCGCRSLUSDGARRSSTTAQPDGRGKAGLGDVHLLMRNTLA
WEICLAPVVKRYITQIRAIQEVETNAALNADVNDTIWAGKILDLPLSRRLNTNA
ATROLLSSSTPFSERRPTTAAPHRPSPSLRSCMEPIPLQAQRQGRQYGLIEK
IGSLGPRHAHTGGDRPVAAVARTSAASRRGRQVSGKHDPFVSHFSKVDSPFAPK
Query Match 16.9%; Score 46.4; DB 8; Length 161266;
Best Local Similarity 52.6%; Pred. No. 6.8;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 74 catggccattccctcccccatttcctgcaccagagaaagcgcattgagcgcgcct 133
Db 35875 CCTAGCTAATCCATCCCCACCACCACCACCACCACCACCACCAGGAGGAC 35816
QY 134 cctccaccctccctccctccctccctccctccctccctccctccctccctcc 193
Db 35815 GCTCGCGCGCGCTCCACCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 35756
QY 194 ctacaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 253
Db 35755 CGACAGCGAGCGCTCCACCCCGCCGCGCCGCCGCCGCCGCCGCCGCC 35696
QY 254 gcaccaagcatgc 265
Db 35695 CGCCGCGATTC 35684
RESULT 11
AE004020 10834 bp DNA BCT 06-JUL-2000
LOCUS xylella fastidiosa, section 166 of 229 of the complete genome.
DEFINITION xylella fastidiosa, section 166 of 229 of the complete genome.
ACCESSION AE004020 AE003849
VERSION AE004020.1 GI:9107122
KEYWORDS xylella fastidiosa.
SOURCE xylella fastidiosa
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
xylella.

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
REMARK  
FEATURES  
source  
gene  
CDS  
gene  
CDS  
gene  
CDS

1. (bases 1 to 10834)  
Silvestri,M.L., Siqueira,W.J., de Souza,A.A., de Souza,A.P.,  
Terenzi,M.F., Truffi,D., Tsai,S.M., Tshako,M.H., Vallada,H., Van  
Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M.,  
Meidanis,J. and Setubal,J.C.  
The genome sequence of the plant pathogen Xylella fastidiosa. The  
Xylella fastidiosa Consortium of the Organization for Nucleotide  
Sequencing and Analysis, Sao Paulo, Brazil  
Nature 406 (6792), 151-157 (2000)  
20365717

2. (bases 1 to 10834)  
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,  
Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,  
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,  
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,  
Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,  
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,  
Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,  
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,  
Frohne,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,  
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,  
Kemper,E.L., Klugajma,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,  
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,  
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,  
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,  
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,  
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nobrega,F.G.,  
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A.A., Nobrega,F.G.,  
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,  
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira  
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,  
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,  
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,M., Silva  
Jr.,W.A., da Silva,J.F., Silvestri,M.L.Z., Siqueira,W.J., de  
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,  
Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,  
Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.  
Direct Submission  
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and  
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP  
13083-970, Brazil

Location/Qualifiers  
1. .10834  
/organism="Xylella fastidiosa"  
/db\_xref="taxon:2371"  
122. .775  
/gene="XF2008"  
122. .775  
/gene="XF2008"  
/note="similar to GI|4981730 (percent identity: 39 %/query  
alignment coverage: 68.7 %/subject alignment coverage:  
83.2 %); identified by sequence similarity; putative; ORF  
located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF84810.1"  
/db\_xref="GI:9107123"  
/translation="MRSMLPSPDFTATLNSRSDPVLFPFHQDFTAMNSNTPRV  
VFEKQRMIVDGVSEYSEYRHPGGLAAIILAVTPERRILFVQFRIPTQRIEMPA  
GLVGIDIDAGEIEQAAIELEFEETGTADHIGLLIGPVSASSERIAFVATGLK  
VSSGGGVDEIVVHEIPRTQAAALVQKMHAGFAMPDKLWALWLDHLDGTPIK"  
complement(899. .2239)  
/gene="XF2009"  
complement(899. .2239)  
/gene="XF2009"  
/note="similar to GI|1653657 (percent identity: 46 %/query  
alignment coverage: 98.9 %/subject alignment coverage:  
100.0 %); identified by sequence similarity; putative; ORF  
located using Glimmer/RBSfinder/Start codon shift: -75"  
/codon\_start=1  
/transl\_table=11

/product="aminopeptidase P"  
/protein\_id="AAF84811.1"  
/db\_xref="GI:9107124"  
/translation="MNRIVKHTGCIAPAEYCRBRRLMKVGPQGIWVLPAAEPRVRS  
RDTHYPRQSDSDFWLCGPEPDAVLVIFGRCGVLLFCRDEPDERAWDQFRAGY  
DAEIQYQMDADYPIDDLDELPLGREGSVYHFGDRVEFDLRLIGWNRVDRQVR  
YAESPHFEIDLGLLHEORLFSYDIIVLLOCAADISVEAHLAALRYAHGVSYYL  
QAELERVRAADSCPAYTSIVAAGANACVLIHYRANACSRDGLIVITDAGAYRGYAA  
DITTFPVNGRFSQAORALVDGAAYNALAOVGPCLPYEAGHAAVQVLTLEGLLRL  
GLLHGTLEDNLADQSYKFEYHKTCHWGLGDVHDVGDYRIDGESRLLEPGMVTIEPG  
LYLVDDTAVHPKRWGIGIRTEDDVLTECHCVLTGALPSADEIAEANAAR"  
complement(2236. .2793)  
/gene="XF2010"  
complement(2236. .2793)  
/gene="XF2010"  
/note="similar to GI|140683 (percent identity: 36 %/query  
alignment coverage: 91.4 %/subject alignment coverage:  
87.1 %); identified by sequence similarity; putative; ORF  
located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF84812.1"  
/db\_xref="GI:9107125"  
/translation="MESPMHLPVIAVQOESQOGLSVTAPELHGSLSGLLAGGGNG  
PDWLAMILADAEVAAPKPSVLSRGLYQATASQLEDPDFAFQLLADGATLAAADAL  
FEWCRFLGFGFLAAHSRSLVSGDEITLRLDLAKLAQASVDDFDMNEEEDGSLEIE  
EFVRVAVLLHGDCLIGCAPQPLN"  
2986. .3207  
/gene="XF2011"  
2986. .3207  
/gene="XF2011"  
/note="hypothetical protein; identified by sequence  
similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF84813.1"  
/db\_xref="GI:9107126"  
/translation="MDQPDALTELHSLAAQLERLEYCQRLAEEHNLRCOOHOMLE  
RELLAKNEQARSVEANVTRIKLTLEQT"  
3204. .3503  
/gene="XF2012"  
3204. .3503  
/gene="XF2012"  
/note="similar to SP|P45580 (percent identity: 27 %/query  
alignment coverage: 91.9 %/subject alignment coverage:  
83.5 %); identified by sequence similarity; putative; ORF  
located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF84814.1"  
/db\_xref="GI:9107127"  
/translation="MSDTEPVSIIRILDREYTVGVGIERKGLIAAQLLDAMHREIRG  
SNRMASIDRIAVLAALNLAHELQQLHEKHADERALLTLDLQQLKANDTSH"  
4005. .4616  
/gene="XF2013"  
4005. .4616  
/gene="XF2013"  
/note="similar to SP|P44905 (percent identity: 35 %/query  
alignment coverage: 88.2 %/subject alignment coverage:  
95.7 %); identified by sequence similarity; putative; ORF  
located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF84815.1"  
/db\_xref="GI:9107128"  
/translation="MLQTSKPMTDNRSLRQHMARRRAIHKHARLAANVTLAEHLIT  
LPFAQTGAVAGYWMADGEIPLHWOTLPQAVRYCLPILAGTKLRFPAHPQGPLKQ  
NRYGPEPDVITDELTPESMTLVVTVPLVGFDTQGRHGMGGGWDRSFAFRNQSP  
PWLGVGVSVQVSTLPTAEWDICVDAICTELTLLPATVST"

```

gene      4637. .5080
          /gene="XF2014"
CDS       4637. .5080
          /gene="XF2014"
          /note="similar to GI:6458261 (percent identity: 53 %/query
          alignment coverage: 98.6 %/subject alignment coverage:
          89.0 %); identified by sequence similarity; putative; ORF
          located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AAF84816.1"
          /db_xref="GI:9107129"
          /translation="MKSEDPAFSIDDLQVRGTEPWNGIRNYQARNFIRDKIHIGDSVL
          FYHNSCKPGVIGLAKIASAYDESDQFNSDYDPKARROPRLVLDIAFERKLA
          RTTLERIKQHALGEGFPLTRGNRLSVFPVTTQWKLILSL"
          5053. .5790
          /gene="XF2015"
          5053. .5790
          /gene="XF2015"
          /note="similar to SP|P27252 (percent identity: 57 %/query
          alignment coverage: 84.5 %/subject alignment coverage:
          94.0 %); identified by sequence similarity; putative; ORF
          located using Glimmer/RBSfinder/Start codon shift: -147"
          /codon_start=1
          /transl_table=11
          /product="ribose-5-phosphate isomerase A"
          /protein_id="AAF84817.1"
          /db_xref="GI:9107130"
          /translation="METATVTVNSTQPNACAPSHPIIINNPMSEAKRRRAEKAIE
          YVNDMIGVGTGVYFIDALGRTPKIKGASVSSEOSTAHLKHGTEVLELHNTG
          TLPYVDGADQDYPKRLIKGGASLTREKIIAEASKOFICIIDPNKQVATLQKPLP
          IEVTPARSLVARQINARTDQGVWREGVITDNGNVILDVHHLRIITDPVKLSQELNQI
          PGVVCGLPARCADVILGSEPPHIL"
          6008. .6223
          /gene="XF2016"
          6008. .6223
          /gene="XF2016"
          /note="hypothetical protein; identified by sequence
          similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84818.1"
          /db_xref="GI:9107131"

Query Match      16.4%; Score 45; DB 1; Length 10834;
Best Local Similarity 70.6%; Pred. No. 23;
Matches 60; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 188  cgcgcctacaagccgctgagtcagctgagtcgcgcgagtcgctcgccctagcaccgg 247
Db 5163  CGCGCGGAAGGCCATCGAATACGTCGGAACGATATGATTCGCGCGCGCACCGG 5222

QY 248  ctcacccgcaagcgcgcgctgcag 272
Db 5223  CTCACAGTCGCTACTTCTATCGAC 5247

RESULT 12
RRPHAS
LOCUS      RRPHAS      3875 bp      DNA      BCT      30-NOV-1993
DEFINITION X66407.1 s44207
ACCESSION  X66407.1 GI:46398
VERSION     X66407.1 GI:46398
KEYWORDS   PHA-synthase.
SOURCE      Rhodococcus ruber.
ORGANISM   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE  1 (bases 1 to 3875)
AUTHORS   Pieper U. and Steinbuechel, A.
TITLE     Direct Submission

```

---

```

JOURNAL    Submitted (25-MAY-1992) A. Steinbuechel, Inst. fur Mikrobiologie
           der Georg-August-Universitaet, Grisebachstr. 8, W-3400 Goettingen,
           FRG
REFERENCE  2 (bases 1 to 3875)
AUTHORS   Pieper U. and Steinbuechel, A.
TITLE     Identification cloning and sequence analysis of the
           poly(3-hydroxyalkanoic acid) synthase gene of the gram-positive
           bacterium Rhodococcus ruber
JOURNAL    FEMS Microbiol. Lett. 96, 73-80 (1992)
REFERENCE  3 (bases 1 to 3875)
AUTHORS   Pieper U. and Steinbuechel, A.
TITLE     Identification, cloning and sequence analysis of the
           poly(3-hydroxyalkanoic acid) synthase gene of the gram-positive
           bacterium Rhodococcus ruber
JOURNAL    FEMS Microbiol. Lett. 75 (1), 73-79 (1992)
MEDLINE    92406022
FEATURES   Location/Qualifiers
           1. .3875
               /organism="Rhodococcus ruber"
               /strain="NCIMB40126"
               /db_xref="taxon:1830"
               83. .598
               /note="ORF1"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA47034.1"
               /db_xref="GI:46399"
               /db_xref="SPTREMBL:Q53048"
               /translation="MRCPLSRTGVSRIIRGCDHPRQSPNAPSARGSGRSRTYADP
               MEDDASRPPTLGELIRQRELAELPMRLAAWVGISNPYLSQIERNLRAPSERVLQAI
               AEQHLSDALUTAEVGRPDGSAVVRAIHEDPLTNAQRSLVEMYEAFFREVTVGKR
               RRGARPDGTE"
               595. .638
               terminator
               644. .698
               terminator
               774. .2462
               gene
               /gene="phbCrr"
               774. .2462
               /gene="phbCrr"
               /codon_start=1
               /transl_table=11
               /product="PHA synthase"
               /protein_id="CAA47035.1"
               /db_xref="GI:581529"
               /db_xref="SPTREMBL:Q53049"
               /translation="MLDBVHKKSLTLDPIGWGPVTSVAGVRNPAQVTAATAEYA
               GRLEIPAAATRVFNANDPDAMPDPRDRFSDTAOWENPAVFSLQSLATRAYVE
               ELTEAGSGDPLQDQKARQANLMDALAPSNFLWNPGLTRAFETGASLLRGARVAA
               HDILNRGLPLKVDSDAFTVCENLAATPGKVVNRNDLIELIQAPQTEQHVHPILAA
               PPWINKYYIIDLAPGRSLAEAWOHGRTVFMVSTYRNDESMRHITMDDYVDGIGTAL
               DVBEITGSKIEVLISICLGGMAAARAFAGDKRVSAFTMLNTLLDYISQVGEIG
               LLTDPATLDLVEFRMRQGGFLSGKEMAGSFDIMRAKDLVFNWVSRMMKGEKPAAFDI
               LANNEDSTSMPEAMHSHYLSRNLAEGLVLDGQPLNLDIADCTVVGCAINDH
               IVPWTSSYQAVNLGGDVRVLTNGGHVAGAVNPGRKRVKFAVGAPDAESGTPPAD
               POWMDEAATRYEHSWEDWTANSNKRAGELVAPPMGSTAHPLEDAPGTYVS"
               2547. .2963
               /note="ORF3"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA47036.1"
               /db_xref="GI:46401"
               /db_xref="SPTREMBL:Q53050"
               /translation="MTTAKTPVDAVAKTTADAAKANEAAKATADATAAOKIAESA
               TAAANQATAEFTKALDDSAARLSSEFNKVVDAAKQSGNLTVDTYRAVSSLLDLQEKL
               ASASPVGWVEDLTAKQVSFARELTATVTTTARDLLK"
               2969. .3009
               terminator
               3041. .>3875
               CDS
               /note="ORF4"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA47037.1"
               /db_xref="GI:46402"
               /db_xref="SPTREMBL:Q53051"
               /translation="MDALQTVATAAPFRIGAAAIAWSSYLATLARRGAHPLDVATD

```









```

/product="Pkbj"
/protein_id="AAF96389.1"
/db_xref="GI:9280391"
/translation="MTAPVKDDPVAHALVAFKTKTRSDWPVDRDLFAEGGLTSLFAM
ELVYLEKTFDVTIAGDPLQLANERTVESMVALVHRLRAVDA"
/gene="fkbk"
/complement(11245..12093)
/gene="fkbk"
/complement(11245..12093)
/gene="fkbk"
/feature="acyl CoA dehydrogenase"
/codon_start=1
/transl_table=11
/product="Pkbk"
/protein_id="AAF96390.1"
/db_xref="GI:9280392"

Query Match      16.3%; Score 44.6; DB 1; Length 77534;
Best Local Similarity 53.1%; Pred. No. 17;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 96 ttatctccaccagaaagccgcatgagccgagccgctctctccaccctctccctccctcc 155
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 46349 TACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 46290
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 156 gtatctctaccaccagacgatttgaagaaataatgccctcctacagccgctcgtacgtg 215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 46289 GCCGAACCTCGCGACTTCTGACGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGAAC 46230
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 216 gagtcggcgatgctctcgctgctgacgacgctgctccaccgccaagtcgctgacgacg 274
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 46229 CACCTGACACAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46171
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AC007789
LOCUS      AC007789      182756 bp      DNA      PLN      03-DEC-1999
DEFINITION Oryza sativa BAC OSJNBa0049B20 genomic sequence, complete sequence.
ACCESSION  AC007789
VERSION    AC007789.1  GI:5042437
KEYWORDS   HTG.
SOURCE     Oryza sativa.
            Oryza sativa.
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
            Buell, R., Benito, M.-I., Lin, X., Mason, T.M., Umayam, L., Shea, T.P.,
            Fujii, C.Y., Shen, M. and Fraser, C.M.
            Oryza sativa BAC OSJNBa0049B20 genomic sequence
            Unpublished
            2 (bases 1 to 182756)
            Benito, M.-I.
            Direct Submission
            Submitted (11-JUN-1999) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            3 (bases 1 to 182756)
            Benito, M.-I.
            Direct Submission
            Submitted (15-JUN-1999) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            4 (bases 1 to 182756)
            Benito, M.-I.
            Direct Submission
            Submitted (18-JUN-1999) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            5 (bases 1 to 182756)
            Benito, M.-I.
            Direct Submission
            Submitted (03-DEC-1999) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            Address all correspondence to:
            Robin Buell or Maria-Ines Benito
            The Institute for Genomic Research
            9712 Medical Center Dr.

```

Rockville, MD 20850, USA  
e-mail:rbuell@tigr.org or mbenito@tigr.org  
BAC clone OSJNBa0049B20 is from *Oryza sativa*.  
The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis and Rice EST databases at TIGR, and the maize EST database at Genbank (<http://www.tigr.org/tdb/at/at.html>, <http://www.tigr.org/tigr.home/tdb/ogi/index.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

## FEATURES

```

source
1..182756
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/clone="OSJNBa0049B20"
/complement(join<179..204,660..775,896..1020,1264..1303,
1587..2987,3531..3610,3722..3766))
/gene="OSJNBa0049B20.1"
/complement(<179..>3766)
/gene="OSJNBa0049B20.1"
/feature="predicted by genscan and genefinder"
/complement(join(179..204,660..775,896..1020,1264..1303,
1587..2987,3531..3610,3722..3766))
/gene="OSJNBa0049B20.1"
/feature="hypothetical protein"
/codon_start=1
/protein_id="AAD38275.1"
/db_xref="GI:5042438"
/translation="MVSSSAATCGPRITAKIDTSVRFGESSINSFAIYTGADGK
LIDGSSPPVHTPIATDHSNAAAAAASVRSKRRMEHSPHLDGAAPKRRDE
RMAAPREATWVILACVPSVSSDGEADGHLAFDWRDPGVSLTLRQSDSVFSP
APRDFCPDRDDHPYVAVDSAGLLLRGARRSAHDFGVALGFDPAFCANGGYIL
CHATLRMAYLYPCSDYERLLIYAGNVGIRRTAADGHPRLIAELQIESGNGTHRT
LLRYSOELGLGMASTRKVPYPPGRRSCGQGVIVHAGLWVLDLSCGLLNCDAFAKP
DMREVPLEGGCKLPYSSDADHAKRCVNSDGLAFVQIHVDYTAAGRAPSTIMISM
WTLQSDAGEESVSLRHRVDEIDHVTYRTKTMPPRRVPLALHPKELGVVFFQ
ITSNSMFAVDLTVTRIVLECKYKMPQLEPMTHSSRWELPHSHICGEDETDG
TSYNVMLYFDHNSLTDELDFNSRDKADELLSTGRLFINPRFQELRNATFPKYL
SFVIVKATDAYEASCLMRDFVSHVSDGNSINVRVSEPYEDDD"
/complement(1959..1980)
/rpt_family="GC-rich"
2892..2946
/rpt_family="CCG)n"
join(<4106..4327,4875..4935,5026..5411)
/gene="OSJNBa0049B20.2"
<4106..>5411
/gene="OSJNBa0049B20.2"
join(4106..4327,4875..4935,5026..5411)
/gene="OSJNBa0049B20.2"
/codon_start=1
/product="putative DNA-binding protein"
/protein_id="AAD38276.1"
/db_xref="GI:5042439"
/translation="MAACFRCAPSVTAASVPAGPSHLPPDGTALSAFVHDVATASTAR

```

```

repeat_region
repeat_region

```

```

mRNA

```

```

gene

```

```

CDS

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:16 ; Search time 320.8 Seconds  
(without alignments)  
320.859 Million cell updates/sec

Title: US-09-300-482-311

Perfect score: 274

Sequence: 1 cttaacattcttctccacc.....gccaaagcgcgcgcgaccg 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	24.8	1118	18 T87859	DNA encoding ribos
2	44.6	16.3	77536	21 A14651	Nucleotide sequenc
3	43	15.7	701	18 T94500	GAL4 DNA binding d
4	43	15.7	38734	20 Z32020	Human METH1 relate
5	42.6	15.5	1371	21 287214	DNA encoding synth
6	42	15.3	771	20 Z12317	Neisseria meningit
7	41.8	15.3	833	15 Q64203	snab gene encoding
8	41.8	15.3	882	19 V64512	M. tuberculosis im
9	41.8	15.3	882	19 V44403	Mycobacterium tube
10	41.8	15.3	882	20 Z19313	M. tuberculosis an
11	41.8	15.3	882	20 Z19101	M. tuberculosis re
12	41.8	15.3	5392	15 Q64201	Sequence comprisin

c 13	41.6	15.2	290	18	T91500	Mycobacterium tube
c 14	41.6	15.2	290	18	T91436	Mycobacterium tube
c 15	41.6	15.2	290	19	V64483	M. tuberculosis im
c 16	41.6	15.2	290	19	V44375	Mycobacterium tube
c 17	41.6	15.2	290	20	Z19285	M. tuberculosis an
c 18	41.6	15.2	290	20	Z19073	M. tuberculosis re
c 19	41.6	15.2	1929	19	V57472	Sorghum bicolor (L
c 20	41	15.0	3632	17	T06978	T. thermophilus ga
c 21	41	15.0	3632	20	X21373	Thermus thermophil
c 22	40.8	14.9	985	19	V64548	M. tuberculosis im
c 23	40.8	14.9	985	19	V44439	Mycobacterium tube
c 24	40.8	14.9	985	20	Z19349	M. tuberculosis an
c 25	40.8	14.9	985	20	Z19137	M. tuberculosis re
c 26	40.4	14.7	1236	16	T45060	Maize T2 cDNA nuc
c 27	40.4	14.7	1248	21	Z87286	S. venezuelae deso
c 28	40.4	14.7	5970	21	Z56003	Contig 002 from co
c 29	40.4	14.7	1244	21	Z87284	S. venezuelae deso
c 30	40.4	14.7	13613	21	Z87319	S. venezuelae deso
c 31	40.2	14.7	13440	19	V52261	Streptococcus pneu
c 32	39.8	14.5	771	20	Z12316	Neisseria meningit
c 33	39.8	14.5	771	20	Z12318	Neisseria gonorrhoe
c 34	39.8	14.5	1574	21	Z48807	Rice inositol 1,3,
c 35	39.8	14.5	2727	6	N50517	Sequence encoding
c 36	39.4	14.4	3331	21	Z99245	Nucleotide sequenc
c 37	39	14.2	2064	14	Q52638	Streptomyces fradi
c 38	39	14.2	2514	17	T30308	Mouse Sox-9 cDNA.
c 39	38.4	14.0	4466	21	A14663	Nucleotide sequenc
c 40	38.4	14.0	4478	21	A14661	Nucleotide sequenc
c 41	38.4	14.0	4547	21	A14664	Nucleotide sequenc
c 42	38.4	14.0	4571	21	A14662	Nucleotide sequenc
c 43	38.4	14.0	77536	21	A14651	Nucleotide sequenc
c 44	38.2	13.9	396	21	Z34971	Wheat beta-caroten
c 45	38.2	13.9	2291	9	N80309	Entire amylase gen

#### ALIGNMENTS

RESULT 1  
T87859  
ID T87859 standard; cDNA; 1118 BP.  
XX  
AC T87859;  
XX  
DT 21-APR-1998 (first entry)  
XX  
DE DNA encoding ribose-5-phosphate isomerase.  
XX  
KW Ribose-5-phosphate isomerase; screen; inhibitor; herbicidal agent;  
KW ss.  
XX  
OS Spinacia oleracea.  
XX  
FH Key Location/Qualifiers  
FT CDS 25..894  
FT /tag= a  
FT /product= ribose-5-phosphate\_isomerase  
XX  
PN WO9737028-A2.  
XX  
PD 09-OCT-1997.  
XX  
PF 26-MAR-1997; 97WO-EP01539.  
XX  
PR 29-MAR-1996; 96DE-4012772.  
XX  
PA (BADI ) BASF AG.  
XX  
PI Kellermann J, Lerchl J, Schmidt R, Schnarrenberger C;  
PI William M;  
XX  
XX WPI; 1997-503113/46.  
DR P-PSDB; W27505.









Sat Nov 4 18:11:26 2000

DE snaB gene encoding enzyme in streptogramin biosynthetic pathway.  
XX Antibiotic; streptogramin; snaA; snaB; snaC; biosynthesis; enzyme;  
KW biosynthetic pathway; Streptomyces pristinaespiralis; ds.  
XX Streptomyces pristinaespiralis.

OS  
XX Key Location/Qualifiers  
FH 1..834  
FT /\*tag= a  
FT

XX FR2696189-A.

XX 01-APR-1994.

XX 25-SEP-1992; 92FR-0011441.

XX 25-SEP-1992; 92FR-0011441.

XX (RHON ) RHONE POULENC RORER SA.

XX Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P;  
PI Thibaut D, Zagorec M;  
XX WPI; 1994-128286/16.  
DR P-PSDB; R54203.

XX DNA involved in streptogramin antibiotic biosynthesis - for  
PT prodn. or bio-conversion of streptogramin(s) or prodn. of  
PT streptogramin intermediates, derivs. or hybrid antibiotics  
XX

XX Claim 2; Page 51-52; 83pp; French.

XX The snaB gene product is involved in the biosynthesis of  
CC streptogramins, antibiotics active against Gram-positive bacteria.  
CC The identification of the sequences encoding the enzymes involved  
CC in the biosynthetic pathway means that they can be isolated and  
CC manipulated. Mutant microorganisms in which a step in the  
CC streptogramin biosynthetic pathway is blocked can be cultured to  
CC produce streptogramin intermediates, which may later be converted  
CC to streptogramin derivatives. Recombinant cells may also be used  
CC for the bioconversion of streptogramins from one form to another or  
CC for the production of hybrid antibiotics.

XX Sequence 833 BP; 106 A; 395 C; 250 G; 82 T; 0 other;

Query Match 15.3%; Score 41.8; DB 15; Length 833;  
Best Local Similarity 52.6%; Pred. No. 0.057;  
Matches 91; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 102 gccacgagaagcgccatgacgcccgcctctccaccctctccctccctccgctc 161

Db 82 gccgagggccgagtgacgcccgtctgatcgacgcccgcgcccgcgcgc 141

QY 162 ctcacccaagagatttgaagaaatcgccctacaagccgctgagtgagtgcc 221

Db 142 caggggcggttcgagacgacgctgacgcccgcgcccgcgctcaccgagcacatc 201

QY 222 ggcagtgctcgccgtaggaccgcccgtccaccgccaagcatgcccgtcgaccg 274

Db 202 ggcctgatcacgcccgcgtcccgccgacgaccaggccctaccacgtgtcccg 254

RESULT 8

V64512

ID V64512 standard; DNA; 862 BP.

XX AC V64512;

XX 27-JAN-1999 (first entry)

XX M. tuberculosis immunogenic polypeptide TbH-29 DNA.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis; ss.  
XX Mycobacterium tuberculosis.

XX WO9816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

XX 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX P-PSDB; WB1726.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis  
XX

XX Example 4; Page 146-147; 230pp; English.

XX This sequence encodes an immunogenic portion of a soluble Mycobacterium  
CC tuberculosis (MT) antigen which can be used in a method for inducing  
CC protective immunity against tuberculosis (TB). This sequence can be  
CC formulated into vaccines and/or pharmaceutical compositions for  
CC immunising against M. tuberculosis infection or may be used for the  
CC diagnosis of tuberculosis.

XX Sequence 882 BP; 141 A; 355 C; 277 G; 109 T; 0 other;

Query Match 15.3%; Score 41.8; DB 19; Length 882;  
Best Local Similarity 51.9%; Pred. No. 0.058;  
Matches 94; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 92 ccatttcacgcccagagaagccgcccattgacgcccgcctctccaccctctctccccc 151

Db 388 ccattcccccacgcccagagcgcgctgaccacgctcgccgagcgcgcccagcca 447

QY 152 ctccgctacatccacccaagacgatttgaagaaatcgccctacaagccgctcgagta 211

Db 448 cgcgcgacgacccgcccgcggtgaccacgcccacacgacgcccgcgacccggtgacca 507

QY 212 cgtggagtcgagtcgctcgcctgacgcccgcgctccacccgcccacgacgctcgca 271

Db 508 cgcgcacacgagcgcgcccgcgacccacgcccgtgaccacgcccacacgacgctcgcccca 567

QY 272 c 272

Db 568 c 568

RESULT 9

V44403

ID V44403 standard; DNA; 882 BP.

XX AC V44403;

XX 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen TbH-29 DNA.

XX Tuberculosis; infection; diagnosis; antigen; TbH-29; ss.

XX Mycobacterium tuberculosis strain H37RV.



```

XX FH Key Location/Qualifiers
XX FT CDS 3..806
XX FT /*tag= a
XX PN WO9816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18214.
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PF (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
XX DR WPI; 1998-251292/22.
XX DR P-PSDB; W64359.
XX XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and diagnosis of tuberculosis
XX PS Claim 4; Page 153-154; 250pp; English.
XX CC This DNA sequence codes for an antigenic portion (see W64359) of
XX CC Mycobacterium tuberculosis antigen TBH-29. It was isolated from
XX CC a M. tuberculosis strain H37Rv genomic library. The invention
XX CC relates to compositions and methods for diagnosing tuberculosis.
XX CC It provides polypeptides (see W64291-W64379) comprising an
XX CC antigenic portion of a soluble M. tuberculosis antigen, or an
XX CC immunogenic portion of an M. tuberculosis antigen, as well as DNA
XX CC sequences encoding such polypeptides, recombinant expression
XX CC vectors and transformed or transfected host cells. Also claimed
XX CC are methods and diagnostic kits for detecting M. tuberculosis
XX CC infection in a patient using these polypeptides, antibodies or
XX CC oligonucleotide probes and primers, for the diagnosis of
XX CC tuberculosis.
XX XX Sequence 882 BP; 141 A; 355 C; 277 G; 109 T; 0 other;

Query Match 15.3%; Score 41.8; DB 19; Length 882;
Best Local Similarity 51.9%; Pred. No. 0.058;
Matches 94; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 92 ccatttcacgccaccgagaaagccgcatggagccggcctctccaccctctctccc 151
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 388 ccattccaccgacgcgcgacgacgcgcggtgacacgctggcgacgacgcgcgacca 447
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 ctccgtcatctccaccgagacgatttgaagaaatcgccgctacaaagccgtcgagta 211
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 448 cgccgcgaccacgcgcggtgaccacgcgcgaagacgcccgcgcgcggtgacca 507
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 cgtgagtcgcgcgtggtctcctggcctaggcaccggtctccaccgccaagcatgcgctcga 271
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 508 cgccgccaacgacgcgcgcgacacgcgcggtgaccacgcgcgcgcgcgcgcgcga 567
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 272 c 272
   |
Db 568 c 568

RESULT 10
Z19313
ID Z19313 standard; DNA; 882 BP.
XX Z19313;
AC Z19313;
XX 05-NOV-1999 (first entry)
DT

```

```

XX DE M. tuberculosis antigen TBH-29 DNA sequence.
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
XX KW immune response; skin test; ss.
XX OS Mycobacterium tuberculosis.
XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US03268.
XX PR 05-MAY-1998; 98US-0072967.
XX PR 18-FEB-1998; 98US-0025197.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
XX DR WPI; 1999-527409/44.
XX DR P-PSDB; Y39156.
XX XX New antigens from Mycobacterium tuberculosis useful in diagnostic
XX PT skin tests and protective or therapeutic vaccines or compositions
XX PS Claim 4; Page 141; 299pp; English.
XX CC The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
XX CC are vaccines and fusion protein containing M. tuberculosis Ag's.
XX CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
XX CC other polypeptides fragments, can be used in pharmaceutical compositions
XX CC or vaccines to generate a protective or therapeutic immune response to
XX CC M. tuberculosis and as reagents in skin tests for diagnosis of
XX CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
XX CC by, T, B or natural killer cells and/or macrophages in
XX CC tuberculosis-immune subjects. Z19249 to Z19460 and Y39083 to Y39225
XX CC are used in the exemplification of the present invention.
XX SQ Sequence 882 BP; 141 A; 355 C; 277 G; 109 T; 0 other;

Query Match 15.3%; Score 41.8; DB 20; Length 882;
Best Local Similarity 51.9%; Pred. No. 0.058;
Matches 94; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 92 ccatttcacgccaccgagaaagccgcatggagccggcctctccaccctctctccc 151
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 388 ccattccaccgacgcgcgacgacgcgcggtgacacgctggcgacgacgcgcgacca 447
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 ctccgtcatctccaccgagacgatttgaagaaatcgccgctacaaagccgtcgagta 211
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 448 cgccgcgaccacgcgcggtgaccacgcgcgaagacgcccgcgcgcggtgacca 507
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 cgtgagtcgcgcgtggtctcctggcctaggcaccggtctccaccgccaagcatgcgctcga 271
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 508 cgccgccaacgacgcgcgcgacacgcgcggtgaccacgcgcgcgcgcgcgcgcga 567
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 272 c 272
   |
Db 568 c 568

RESULT 11
Z19101
ID Z19101 standard; DNA; 882 BP.
XX Z19101;
AC Z19101;
XX

```



PD 13-MAR-1997.  
XX  
XX 30-AUG-1996; 96WO-US14674.  
XX  
PR 12-JUL-1996; 96US-0680574.  
PR 01-SEP-1995; 95US-0523436.  
PR 22-SEP-1995; 95US-0533634.  
PR 22-MAR-1996; 96US-0620874.  
PR 05-JUN-1996; 96US-0659683.  
XX  
XX (CORI-) CORIXA CORP.

PA Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
XX Twardzik DR, Vedvick TH;  
PI WPI; 1997-192903/17.  
XX

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
PT useful in vaccines for prevention or treatment of tuberculosis, also  
PT for diagnosis

PS Claim 4; Page 78; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence encodes a specifically claimed  
CC M.tuberculosis antigen, TBW-7. The immunogenic protein, and fusion  
CC proteins containing one or more of the proteins or one of the proteins  
CC plus ESAT-6, are useful in vaccines, preferably when formulated with a  
CC non-specific adjuvant, to induce an immune response against  
CC M.tuberculosis (for treatment or prevention).

XX Sequence 290 BP; 30 A; 83 C; 142 G; 33 T; 2 other;

Query Match 15.2%; Score 41.6; DB 18; Length 290;  
Best Local Similarity 51.6%; Pred. No. 0.052;  
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 89 cccccatttcacccacgagaaagccgctgagcgcgcctctccaccctctc 148  
DB 228 CGCCTCGCCACCCCTTACCCCGCTTCCACCGAAGCGCGCTCCGGGCGTTGCCCTCCGC 169  
QY 149 cccctccgtcatctcaccacgaagcatttgaagaaatcgccgctacaaagccgtga 208  
DB 168 CACCGCCGCGCGCCCAAGACCGCGCTTGCCTGGAGCGCGCCATTGCCGCCCT 109  
QY 209 gtacgtgagtcgcgcgcctgctcgccttaggcacccgctccaccgcaagcgtgcgt 268  
DB 108 GCCCACCAGGCGCGCTGGCGCGCGCACCGCAAGACGCGCTGCGCGCGCGCGC 49  
QY 269 cgac 272  
DB 48 CGAC 45

RESULT 14  
T91436/C  
ID T91436 standard; DNA; 290 BP.  
XX  
XX T91436;  
AC

DT 13-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen TBW-7 encoding DNA.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis; ss.

XX Mycobacterium tuberculosis.

XX WO9709429-A2.

XX 13-MAR-1997.  
XX  
XX 30-AUG-1996; 96WO-US14675.  
XX  
PR 12-JUL-1996; 96US-0680573.  
PR 01-SEP-1995; 95US-0523435.  
PR 22-SEP-1995; 95US-0532136.  
PR 22-MAR-1996; 96US-0620280.  
PR 05-JUN-1996; 96US-0658800.  
XX

PA (CORI-) CORIXA CORP.  
XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
XX Twardzik DR, Vedvick TH;  
PI WPI; 1997-192904/17.  
XX

DR New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
XX - useful for diagnosis of M. tuberculosis infection  
PT  
PT  
PS Claim 4; Page 82-83; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence encodes a specifically claimed  
CC M.tuberculosis antigen, TBW-7. The immunogenic polypeptide can be  
CC used to diagnose M.tuberculosis infection by forming complexes with  
CC specific antibodies in the sample. Fragments of DNA encoding the  
CC immunogenic polypeptide can be used as diagnostic primers or probes  
CC and agents that bind to the antigen, especially monoclonal antibodies  
CC or equivalent polyclonal antibodies, are also used for diagnosis.

XX Sequence 290 BP; 30 A; 83 C; 142 G; 33 T; 2 other;

Query Match 15.2%; Score 41.6; DB 18; Length 290;  
Best Local Similarity 51.6%; Pred. No. 0.052;  
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 89 cccccatttcacccacgagaaagccgctgagcgcgcctctccaccctctc 148  
DB 228 CGCCTCGCCACCCCTTACCCCGCTTCCACCGAAGCGCGCTCCGGGCGTTGCCCTCCGC 169  
QY 149 cccctccgtcatctcaccacgaagcatttgaagaaatcgccgctacaaagccgtga 208  
DB 168 CACCGCCGCGCGCCCAAGACCGCGCTTGCCTGGAGCGCGCCATTGCCGCCCT 109  
QY 209 gtacgtgagtcgcgcgcctgctcgccttaggcacccgctccaccgcaagcgtgcgt 268  
DB 108 GCCCACCAGGCGCGCTGGCGCGCGCACCGCAAGACGCGCTGCGCGCGCGCGC 49  
QY 269 cgac 272  
DB 48 CGAC 45

RESULT 15  
V64483/C  
ID V64483 standard; DNA; 290 BP.  
XX  
XX V64483;  
AC

DT 27-JAN-1999 (first entry)

XX M. tuberculosis immunogenic polypeptide TBW-7 DNA.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis; ss.

XX Mycobacterium tuberculosis.

PN	W09816646-A2.		
XX			
XX	23-APR-1998.		
PD			
XX			
XX	07-OCT-1997; 97WO-US18293.		
PE			
XX			
XX	13-MAR-1997; 97US-0818112.		
PR			
PR	11-OCT-1996; 96US-0730510.		
XX			
XX	(CORI-) CORIXA CORP.		
PA			
XX			
XX	Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;		
PI	Reed SG, Sheikly YAW, Twardzik DR, Vedwick TS;		
PI			
XX			
XX	WPI; 1998-261042/23.		
DR			
XX			
XX	Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used		
PT	to develop products for the detection of M. tuberculosis infection		
PT	and for diagnosis, treatment and prevention of tuberculosis		
PT			
XX			
PS	Claim 4; Page 84-85; 230pp; English.		
PS			
XX			
CC	This sequence encodes an immunogenic portion of a soluble Mycobacterium		
CC	tuberculosis (MT) antigen which can be used in a method for inducing		
CC	protective immunity against tuberculosis (TB). This sequence can be		
CC	formulated into vaccines and/or pharmaceutical compositions for		
CC	immunising against M. tuberculosis infection or may be used for the		
CC	diagnosis of tuberculosis.		
XX			
XX	Sequence 290 BP; 30 A; 83 C; 142 G; 33 T; 2 other;		
SQ			
Query Match		15.2%;	Score 41.6; DB 19; Length 290;
Best Local Similarity		51.6%;	Pred. No. 0.052;
Matches		95; Conservative	0; Mismatches 89; Indels 0; Gaps 0;
Qy	89 cccccatttcacccaccggaagcgcgcgatggacgcggctctctccacctctc	148	
Db	228 CGCCTGGGCCACCTTACC GCCGTGTGCGCACGAAGACCGCCGTCGGGGCGGTGCCTCGCG	169	
Qy	149 cccttcgcgtcatccaccgaagcagttagaataaacgcccgctacaaggcgcgtca	208	
Db	168 CACGCGCCGCGCGCCCAAGACCGCGGTTCGCGCGGTGGAGCCGCGCATTCGCGCCT	109	
Qy	209 gtactgtgagtcgcgcatggtctcgttgctaggcacgcggtctccaccgccaagcatgcgt	268	
Db	108 GCCCACCAGGCGCGCCTTGCCGCGCGCACCGGCAAGAACACGCGCTGCGCCGCCGCCGCGC	49	
Qy	269 cgac 272		
Db	48 CGAC 45		

Search completed: November 4, 2000, 13:45:31  
Job time: 16458 sec











```
Query Match
Best Local Similarity 14.7%; Score 40.4; DB 2; Length 1236;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 97 tcatgcaccagagaaagccgccttgacgcccgtctctccaccctctccctcccg 156
Db 655 TCTCGTCTGCGAGGCTGCGGCGCTGCTGGCGGCTCGCGGCGCTACACGCGCT 714

QY 157 tcatctcaccacagagattgaagaaatcgcgcctcacaagcgctcgatcg 216
Db 715 CCAAGCAGCCATCTGCGGCTACCAAGACGCGCTCGAGCTGCGGCGCAGGGG 774

QY 217 agtcggcgatgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 274
Db 775 TCCGGTCAACTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 832

RESULT 8
US-09-320-878-21/c
; Sequence 21, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-21

Query Match
Best Local Similarity 14.7%; Score 40.4; DB 5; Length 5970;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 112 aagcgcgaatgacgcccgtctcctccaccctcctccctccgtcctcaccga 171
Db 484 ACGCCGCGACGCCCTCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 425

QY 172 acgatttgaagaaatcgcgcctcacaagcgctcgatcgatcgatcgatcg 231
Db 424 CCGAGGTCTTACGCTTACAGCCACCAAGCGCGCTGACAGCGCTGACAGCG 365

QY 232 tcggcctagcagcggctcaccacccaagcatgctcgacc 273
Db 364 TCACGAGAGAGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 323

RESULT 9
US-08-343-428-1
; Sequence 1, Application US/08343428
; Patent No. 5665586
```

```
GENERAL INFORMATION:
APPLICANT: Nakamura, Etsuo
APPLICANT: Tsuzuki, Hiroshige
APPLICANT: Kitadokoro, Kengo
APPLICANT: Shin, Nasaru
APPLICANT: Teraoka, Hiroshi
TITLE OF INVENTION: No. 5665586el Protease
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: James F. Haley, Jr., Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS Dos 5.0
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,428
FILING DATE: 18-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/J93/00592
FILING DATE: 30-APR-1993
APPLICATION NUMBER: JAPAN 4-126511
FILING DATE: 19-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Haley, Jr., James F.
REGISTRATION NUMBER: 27794
REFERENCE/DOCKET NUMBER: SHGN-7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)596-9000
TELEFAX: (212)596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2064
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces fradiae
FEATURE:
NAME/KEY: -35 signal
LOCATION: 359..364
IDENTIFICATION METHOD: by experiment
NAME/KEY: -10 signal
LOCATION: 378..383
IDENTIFICATION METHOD: by experiment
NAME/KEY: CDS
LOCATION: 435..1505
IDENTIFICATION METHOD: by experiment
NAME/KEY: sig peptide
LOCATION: 435..944
IDENTIFICATION METHOD: by experiment
US-08-343-428-1
```

```
Query Match
Best Local Similarity 14.2%; Score 39; DB 1; Length 2064;
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 137 ccaccctctctccctccgtctcctcctcctcctcctcctcctcctcctcctc 196
Db 557 CGACTCCGCGCGCGCGCTCCACGCGCTCGAGCGCGCTCGAGCGCGCTCGGCG 616

QY 197 caagcgctcagtagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 256
Db 617 CAGCGCGCGCGCGCTACGTGGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 676
```

us-09-300-482-311.rni

Sat Nov 4 18:11:26 2000

```
QY 257 caagcatgccgtcga 271
Db 677 GGCCGCGCGGCCAA 691

RESULT 10
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTgpt-F1s
US-08-232-463-14

Query Match 14.2%; Score 39; DB 1; Length 7218;
Best Local Similarity 3.2%; Pred. No. 0.17;
Matches 6; Conservative 118; Mismatches 63; Indels 0; Gaps 0;

QY 1 ctatattcttctccacacacatttgaacctctgggactggtagcttcaa 60
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1257 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1316
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 cctttaacattaacatggccattccctacccttcatgccaccgagaaagcgcga 120
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1317 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1376
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 tggacgcggcctctccaccctctccctccgtctatctccacccacagattga 180
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1377 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1436
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 agaaat 187
; | | | | |

Db 1437 ACCRAAT 1443

RESULT 11
US-08-042-747A-7
; Sequence 7, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042,747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2943 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 87..2744
US-08-042-747A-7

Query Match 13.7%; Score 37.6; DB 1; Length 2943;
Best Local Similarity 49.5%; Pred. No. 0.31;
Matches 97; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 79 ccatttcctacccccatttccatccacccagacatttgaagaaatcgccgctcaca 138
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1102 CCAAGCGCCCTCGGTGTCGACGCTGACCAAGTGGCGGGAGGTGGACGAGATGCTCGCGG 1161
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 139 accctctccctccgtctcatctcaccacacacatttgaagaaatcgccgctcaca 198
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1162 CCGAGTAGGCGCCCTCGTTCGGCTTCTCTCGCGCCCTCTCGACACACCTTCACCGCCA 1221
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 199 aggcgcgtcgagtgcgttcggcgatggtctctggtcgttaggcacgcggtccacgcga 258
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1222 ACCGACACGAGTAGCGCTCTCGCGGCTCGACCTCGCGGACTCGTCGGGCGCGAGGCC 1281
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 259 agcatgcgctgaccg 274
; | | | | |
Db 1282 GCGAGGCGGTGGACCG 1297
```



QY 243 accggtccaccgccaagcatgccgtc 269  
||||| | | | | | | |  
db 26959 ACCGCCGGGCACTCGCCGAGGTCGAC 26985

Search completed: November 4, 2000, 13:34:08  
Job time: 16684 sec

RESULT 15  
US-08-804-227C-7  
; Sequence 7, Application US/0804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rosteck, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:

CORRESPONDENCE ADDRESS: 1501  
 ADDRESSEE: THOMAS G. PLANT  
 STREET: LILLY CORPORATE CENTER  
 CITY: INDIANAPOLIS  
 STATE: IN  
 COUNTRY: USA  
 ZIP: 46285

```

? ZIP: 48263
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: ASCII(DOS) Text only
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/804,227C
? FILING DATE: February 21, 1997
? CLASSIFICATION: 435
?
```

CLASSIFICATION: 455  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:

```

: INFORMATION FOR SEQUENCE ANALYSIS:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 44377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single

```

```
; STRANDEDNESS: single  
; TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)
```

```

; MOLECULE FILE:  SMN (3)
;
; FEATURE:
; NAME/KEY:  CDS

```

```

; NAME/REF: CSC
; LOCATION: 350..14002
; FEATURE:

```

NAME/KEY:	CDS
LOCATION:	14046..20036

```

FEATURE:
NAME/KEY: CDS

```

```

; LOCATION: 20110..31284
; FEATURE:

```

```
NAME/KEY: CDS  
LOCATION: 31329..36071
```

```

; ;
; ;
; ; NAME/KEY: CDS

```

LOCATION:  
US-08-804-227C-7

Query Match	13.4%;	Score 36.6;	DB 3;	Length 44377;
Best Local Similarity	53.1%;	Pred. No. 1;		
Matches 78: Conservative	0;	Mismatches 69;	Indels 0;	Gaps 0;

[illegible][illegible][illegible][illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:13 ; Search time 4352.3 Seconds  
(without alignments)  
389.240 Million cell updates/sec

Title: US-09-300-482-311

Perfect score: 274

Sequence: 1 ctacattcttcttccacc.....gccaaagcatgcgtgacccg 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estpl1.\*  
71: em\_estpl2.\*  
72: em\_estpl3.\*  
73: em\_estpl4.\*  
74: em\_estpl5.\*  
75: em\_estro1.\*  
76: em\_estro2.\*  
77: em\_estro3.\*  
78: em\_estro4.\*  
79: em\_estro5.\*  
80: em\_estro6.\*  
81: em\_estro7.\*  
82: em\_estro8.\*  
83: em\_estro9.\*  
84: em\_estro10.\*  
85: em\_estro11.\*  
86: em\_estro12.\*  
87: em\_estro13.\*  
88: gb\_gss1.\*  
89: gb\_gss2.\*  
90: gb\_gss3.\*  
91: gb\_gss4.\*  
92: em\_gss1.\*  
93: em\_gss2.\*  
94: em\_gss3.\*  
95: em\_gss4.\*  
96: gb\_gss5.\*  
97: gb\_gss6.\*  
98: gb\_gss7.\*  
99: gb\_gss8.\*  
100: gb\_gss9.\*  
101: em\_gss5.\*  
102: em\_gss6.\*  
103: em\_gss7.\*  
104: em\_gss8.\*  
105: em\_gss9.\*  
106: em\_gss10.\*  
107: em\_gss11.\*  
108: gb\_gss10.\*  
109: gb\_gss11.\*  
110: em\_gss12.\*  
111: gb\_gss12.\*  
112: gb\_gss13.\*  
113: gb\_gss14.\*  
114: gb\_gss15.\*  
115: gb\_gss16.\*  
116: gb\_gss17.\*









```

/note=Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI; isolate=Miyakojima MG-20"
BASE COUNT      54 a      94 c      30 g      88 t
ORIGIN
Query Match      36.1%; Score 99; DB 18; Length 266;
Best Local Similarity 78.7%; Pred. No. 1.9e-16;
Matches 133; Conservative 0; Mismatches 30; Indels 6; Gaps 1;
Qy 57 tcaaccttaacattaacatgccattccctaccgccatttcacgcacgagaaagcc 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
bb 104 tctctccctccctcttctctctctctctctctctctctctctctctctctctctct

```

117 gccatgacgcggcctctctcaacccctctctcccccctcgtcatactctcacccaagacgat 176  
Qy  
118  
Db  
164 gccattgaagcaggtcttcgcaccat-----cctccatcatctcaccgaagacgat 217  
Qy  
177 ttgaagaaatcgcgcctacaagcgcgtcagtcagtcggagtcgcggca 225  
Qy  
218 ttgaagaaatcgcgcgcctacaaagccgcgtcagtcagtcgaaatccggca 266  
Db

[illegible]

```

ADDRESS=
COMMENT=
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers
1. .309
/organism="Lotus japonicus"
/db_xref="taxon:34305"
/clone="WM131e10.1"
/clone_lib="Lotus japonicus young plants (two-week old)"
/dev_stage="young plants (two-week old)"
/notes="Vector: pBlueScriptII SK-; Site_1: EcorI; Site_2:
XhoI; isolate=Miyakojima MG-20"
BASE COUNT 63 a 130 c 48 g 68 t
ORIGIN
Query Match 32.1%; Score 88; DB 18; Length 309;
Best Local Similarity 80.5%; Pred. No. 1.6e-13;
Matches 103; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```

QY	205	tcagtagctaggagtcggaatggtctcgtcgctagcaccggtctccaccgccaagcatg	264
Db	238	TCGACTACGTCAAGTCGGGATGGTCTCGGCTCGGCACCGGCTCCACGCCGCTTCG	297
QY	265	cagtcgac	272



	Query Match	32.1%;	Score 88;	DB 18;	Length 382;
	Best Local Similarity	80.5%;	Pred. No. 1.6e-13;		
	Matches 103; Conservative	0;	Mismatches 25;	Indels 0;	Gaps 0;
YY	145	cctcccccttcggtcattctcaaccacagcgatttgaagaaaaatgcggcgtacaaggccg	204		
bB	210	CCATTCGTCGCATCACCTTCACCCAAGACGACCTCAAGAGACTCGCGCGCGACAAGGCCG	269		
YY	205	tctagtlactgttgagtccggcatggtccctcgcccttagcacccggtctcaaccgccaaacatg	264		
bB	270	TGGACTACGTCAAGTCCGGCATGTCTTCGGCCTCGGCGACCGGGCTCCACCGCCGCTTCG	329		
YY	265	cogtgcgac	272		
bB	330	TCGTGCCC	337		

RESULT	14
AV407718	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	

[illegible]

AV407718 399 bp mRNA EST 23-MAY-2000  
AV407718 Lotus japonicus young plants (two-week old) Lotus  
japonicus cDNA clone MWL029c05\_I\_5', mRNA sequence.  
AV407718  
AV407718.1 GI:7720572  
EST.  
Lotus japonicus.  
Lotus japonicus.  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
Fabales; Fabaceae; Papilionoideae; Lotus.  
1 (bases 1 to 389)  
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
Generation of 7137 non-redundant expressed sequence tags from a  
legume, Lotus japonicus  
DNA Res. 7 (2), 127-130 (2000)

```

Query Match      32.1%; Score 88; DB 18; Length 395;
Best Local Similarity 80.5%; Pred. No. 1.6e-13;
Matches 103; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

145 cctcccctcgcgtcatctccaccacagacgatttgaagaaatccgcgcgtactacaaggcgg 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 CCATCCGCTGCCCATCACCTCACCACAGACACACTCAAGAGACTCGCCGCCGCACAAGCGCG 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 tcgagtacgtgagtcgcgcatggtcctcggcctagcagccgcgtccaccgccaaagcatg 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 TCGACTAGTGTCAAGTCCGGCATGCTCTCGGCCTCGGCACCGCGCTCCACCGCGCGCTTCG 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY          265 cegtcgac 272
                ||||| |
Db          309 TCGTCCGC 316

```

Sat Nov 4 18:11:28 2000

**MEDLINE COMMENT**  
20277479  
Contact: Yasukazu Nakamura  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: ynakamu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.  
Location/Qualifiers

	Query Match	32.1%	Score 88;	DB 18;	Length 389;
	Best Local Similarity	80.5%;	Pred. No. 1.6e-13;		
	Matches 103;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;
QY	145	cctcccccctcgctcatcctcaccaagacatttgagaaaatcgcgcgtacaaggccg	204		
Db	179	CCATCGGTCATCACCTCACCACAGCAGACTCAAGAGACTCGGCCGCCGAAGGC	238		
QY	205	tcgagtactgtagtccggcgatgctctcgcttagtcaggaccgcgtccaccgccaaagcatg	264		
Db	239	TGCACTAGCTCAAGTCCGGCATGTCTCGGCTTCGGCACCGGCTCCACCGCGCTTCG	298		
QY	265	ccgtgcag	272		
Db	299	TCGTGCC	306		

RESULT	15
AV413372	
LOCUS	395 bp mRNA EST 23-MAY-2000
DEFINITION	AV413372 Lotus japonicus young plants (two-week old) Lotus japonicus cDNA clone MWM231a10_r 5', mRNA sequence.
ACCESSION	AV413372
VERSION	AV413372.1 GT:7742548
KEYWORDS	EST.
SOURCE	Lotus japonicus.
ORGANISM	Lotus japonicus Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus. 1 (bases 1 to 395) Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S. Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus DNA Res. 7 (2), 127-130 (2000)
REFERENCE	20277479
AUTHORS	Contact: Yasukazu Nakamura
TITLE	The First Laboratory for Plant Gene Research
JOURNAL	Kazusa DNA Research Institute
MEDLINE	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
COMMENT	Email: vnakamu@kazusa.or.jp, URL: <a href="http://www.kazusa.or.jp/en/plant/">http://www.kazusa.or.jp/en/plant/</a>

```

FEATURES
  source
    Location/Qualifiers
      .395
      /organism="Lotus japonicus"
      /db_xref="taxon:34305"
      /clone="MM231a10.1"
      /clone_lib="Lotus japonicus young plants (two-week old)"
      /dev_stage="young plants (two-week old)"
      /note="Vector: pBluescript1 SK-; Site_1: EcoRI; Site_2:
      xhoI; isolate=Miyakojima MG-20"
82 a 162 c 67 g 84 t
BASE COUNT
ORIGIN

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:34:27 ; Search time 4075.18 seconds  
(without alignments)  
415.840 Million cell updates/sec

Title: US-09-300-482-298  
Perfect score: 388

Sequence: 1 ggagaaagaagaagaagatg.....ttaagcctgggagcccccgt 388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pr1: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: em\_fun: \*  
13: em\_hum1: \*  
14: em\_hum2: \*  
15: em\_in: \*  
16: em\_om: \*  
17: em\_or: \*  
18: em\_ov: \*  
19: em\_ph: \*  
20: em\_pl: \*  
21: em\_ro: \*  
22: em\_sts: \*  
23: em\_sy: \*  
24: em\_un: \*  
25: em\_vl: \*  
26: gb\_htg1: \*  
27: gb\_htg2: \*  
28: gb\_in1: \*  
29: gb\_in2: \*  
30: gb\_ba1: \*  
31: em\_ba2: \*  
32: em\_hum3: \*  
33: gb\_pr4: \*  
34: gb\_htg3: \*  
35: gb\_htg4: \*  
36: gb\_htg5: \*  
37: gb\_htg6: \*  
38: gb\_htg7: \*  
39: em\_htg1: \*  
40: em\_htg2: \*  
41: em\_htg3: \*

44: em\_hum5: \*  
45: gb\_pl3: \*  
46: gb\_pr5: \*  
47: gb\_htg8: \*  
48: gb\_htg9: \*  
49: gb\_htg10: \*  
50: gb\_htg11: \*  
51: gb\_htg12: \*  
52: gb\_htg13: \*  
53: gb\_htg14: \*  
54: gb\_in3: \*  
55: gb\_htg15: \*  
56: gb\_htg16: \*  
57: gb\_htg17: \*  
58: em\_htg4: \*  
59: em\_htg5: \*  
60: em\_htg6: \*  
61: em\_htg7: \*  
62: em\_hum6: \*  
63: gb\_htg18: \*  
64: gb\_htg19: \*  
65: gb\_ba3: \*  
66: em\_htg8: \*  
67: em\_htg9: \*  
68: em\_htg10: \*  
69: em\_htg11: \*  
70: em\_htg12: \*  
71: em\_htg13: \*  
72: em\_htg14: \*  
73: em\_htg15: \*  
74: em\_htg16: \*  
75: em\_htg17: \*  
76: em\_htg18: \*  
77: em\_htg19: \*  
78: em\_htg20: \*  
79: em\_htg21: \*  
80: em\_htg22: \*  
81: em\_htg23: \*  
82: gb\_pr6: \*  
83: gb\_pr7: \*  
84: gb\_htg20: \*  
85: gb\_htg21: \*  
86: gb\_htg22: \*  
87: gb\_htg23: \*  
88: gb\_ro: \*  
89: gb\_sts1: \*  
90: gb\_sts2: \*  
91: gb\_sy: \*  
92: gb\_un: \*  
93: gb\_vil: \*  
94: gb\_vil2: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
1	239.8	61.8	930	7	AF189365	AF189365	Oryza sat
2	85.2	22.0	14229	53	AC057311	AC057311	Giardia
3	85.2	22.0	14229	53	AE001286	AE001286	Chlamydia
4	80.8	20.8	122925	85	AL360001	AL360001	Homo sapi
5	79.2	20.4	153804	27	AC006280	AC006280	Plasmodiu
6	79.2	20.4	196149	27	AC004709	AC004709	Plasmodiu
7	79.2	20.4	911	45	SCPOS18	SCPOS18	S.cerevisia
8	79.2	20.4	1328	45	SCYJL121C	SCYJL121C	S.cerevisia
9	78.6	20.3	5960	2	RCU23145	RCU23145	Rhodobacter
10	77.6	20.0	12939	1	AE002306	AE002306	Chlamydia
11	76.4	19.7	10150	1	AE002217	AE002217	Chlamydia
12	76.4	19.7	12973	1	AE001604	AE001604	Chlamydia

AF002545 Chlamydomonas reinhardtii  
 AF070943 Expressio  
 AF070942 Expressio  
 L42328 Spinacia ol  
 AF070941 Spinacia  
 AC022355 Arabidops  
 U97573 Treponema p  
 AF001263 Treponema  
 Z98979 S.pombe chr  
 Z50098 S.tuberosum  
 D90911 Synchocyst  
 AC010797 Arabidops  
 AF047444 Oryza sat  
 Y13937 Bacillus su  
 Z99112 Bacillus su  
 AF269471 Staphyloc  
 AF270217 Staphyloc  
 AF270200 Staphyloc  
 AE000482 Escherich  
 Continuation (2 of  
 AC034250 Homo sapi  
 U00006 E. coli chr  
 AF015274 Arabidops  
 AE001811 Thermotog  
 AC019942 Drosophil  
 U28991 Caenorhabdi  
 AC005641 Drosophil  
 AC005450 Drosophil  
 AE003840 Drosophil  
 Z80108 Mycobacteri  
 AE003874 Xylella f  
 Z78161 M.musculus  
 AE004329 Vibrio ch

76.4 19.7 300050 1 AP002545  
 75.4 19.4 849 91 AF070943  
 75.4 19.4 993 91 AF070942  
 75.4 19.4 1198 45 SPIR5P3E  
 75.4 19.4 1230 7 AF070941  
 75.4 19.3 67712 8 AFAC022355  
 74.2 19.1 3400 65 TPUS97573  
 74.2 19.1 10081 1 AE001263  
 74.2 19.1 39729 45 SPAC31G5  
 74.2 19.1 1055 45 STPEPIMR  
 72.6 18.7 143051 2 D90911  
 71.4 18.4 89154 8 AFAC010797  
 69 17.8 1050 7 AF047444  
 68.6 17.7 27779 2 BS113937  
 68.6 17.7 208780 2 BSUB0009  
 68 17.5 3246 1 AF270217  
 68 17.5 3276 1 AF270200  
 66.4 17.1 20906 1 AE000482  
 66.4 17.1 110000 28 AC008693\_1  
 66.4 17.1 110000 52 AC034250\_0  
 66.4 17.1 176195 2 ECUW89  
 65.8 17.0 1259 7 AF015274  
 65.4 16.9 10206 1 AE001811  
 63.2 16.3 26292 40 AC019942  
 63.2 16.3 27521 30 CELF08F8  
 63.2 16.3 50089 29 AC005641  
 63.2 16.3 125150 27 AC005450  
 63.2 16.3 259764 29 AE003840  
 60.6 15.6 39150 2 MTCY21B4  
 60.2 15.5 10743 1 AE003874  
 59.2 15.3 435 8 MMZ78161  
 59 15.2 12506 1 AE004329

RESULT 1  
 AF189365  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 Source  
 CDS

AF189365  
 Oryza sativa  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
 1 (bases 1 to 930)  
 Kopriova, S., Kopriova, A. and Suess, K.H.  
 Identification, cloning, and properties of cytosolic  
 D-ribulose-5-phosphate 3-epimerase from higher plants  
 J. Biol. Chem. 275 (2), 1294-1299 (2000)  
 20092904  
 2 (bases 1 to 930)  
 Kopriova, S., Kopriova, A. and Suess, K.H.  
 Direct Submission  
 Submitted (20-SEP-1999) Institute of Forest Biology and Tree  
 Physiology, Am Flughafen 17, Freiburg im Breisgau 79085, Germany  
 Location/Qualifiers  
 1. 930  
 /organism="Oryza sativa"  
 /db\_xref="taxon:4530"  
 35..721  
 /EC\_number="5.1.3.1"  
 /note="cytosolic isoform"  
 /codon\_start=1  
 /product="D-ribulose-5-phosphate 3-epimerase"  
 /protein\_id="AA01048.1"  
 /db\_xref="GI:6007803"  
 /translation="MAAAAKIAPSMILSSDFANLAAEADRMVRLGADWLHMDIMGH  
 FVPNLITGAPVIOSLRKHKAYLIDCHLWNTNPSDYVEFLAKAGASGFTFIEVSRDNW

Query Match 61.8%; Score 239.8; DB 7; Length 930;  
 Best Local Similarity 78.8%; Pred. No. 66-55; 77; Indels 0; Gaps 0;  
 Matches 286; Conservative 0; Mismatches 77;

QY 26 gacacgcaaaatagctcccttcgctgctctctccgacttcgcaatttgcttcgagggc 85  
 DB 49 GCGCGCAAGATAGCGCGCTGCTCTCGTGGACTTCGCCAACCTCGCGCGGAGGC 108  
 QY 86 tcagcgcgactcccaacttcgcccgcgcttcgctcccaatgagacatcatgagtcattt 145  
 DB 109 CGACCGCATGCTCGGCTCGCGCCGACTGGCTCCCATGGACATCATGGACGGGCAT 168  
 QY 146 tgcccccaatttaactattgctccgctccagttattgaaagtttgagaaagcacacaaagc 205  
 DB 169 TGTTCCTAATCTTACATATTGAGCTCCAGTGATTCAGAGCTTCAGGAAGCACACCAAGGC 228  
 QY 206 atattggattgacaccttatggtacacaaatctctctgattgattgtaaccttgcaaa 265  
 DB 229 ATATTGGACTGCCATCTTATGTTGACCAATCCTTCGGATTATGTAGAACATTAGCAA 288  
 QY 266 agctggcttcgctgcttcacatttcacatgtagagacataaaagataaactggaaagact 325  
 DB 289 AGCTGGTCCCTCAGGTTTCACATTCATCCATATAGAGTATCCAGAGACATATGGCAGAACT 348  
 QY 326 tatccaaagaatcaagtcacatgcgcatgattctctggtgtagcattaaagcctgggacccc 385  
 DB 349 CATCAAAGTATCAAGCAAGAGGTATGCGACCGGGTGTATCATTTGAGCGCAGGCATCC 408  
 QY 386 cgt 388  
 DB 409 TGT 411

RESULT 2  
 AC057311  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 Source  
 CDS

AC057311  
 Giardia intestinalis clone MU2761 strain WB-C6, LOW-PASS SEQUENCE  
 SAMPLING.  
 AC057311  
 GI:7594009  
 HTG: HTGS PHASE0.  
 Giardia intestinalis.  
 Giardia intestinalis.  
 Eukaryota; Diplomonadida; Hexamitidae; Giardia.  
 1 (bases 1 to 927)  
 Morrison, H.G., McArthur, A.G., Nixon, J., Eakin, N.O., Kim, U.,  
 Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.  
 Giardia: a model for ancient eukaryotic genome analysis  
 Unpublished  
 2 (bases 1 to 927)  
 Morrison, H.G., McArthur, A.G., Nixon, J., Eakin, N.O., Kim, U.,  
 Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.  
 Direct Submission  
 Submitted (18-APR-2000) Josephine Bay Paul Center for Comparative  
 Molecular Biology and Evolution, Marine Biological Laboratory, 7  
 MBL Street, Woods Hole, MA 02543-1015, USA  
 \* NOTE: This record contains 1 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will







Contact: hunquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: BA332019  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 103697 bases at least Q40  
 Consensus quality: 110797 bases at least Q40  
 Consensus quality: 115383 bases at least Q30  
 Insert size: 120025; sum-of-contigs  
 Insert size: 170370; 12.7% error; agarose-fp  
 Quality coverage: 0.00x in Q20 bases; sum-of-contigs Quality  
 Coverage: 0.00x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 3339: contig of 3339 bp in length
* 3340 3439: gap of 100 bp
* 3440 6321: contig of 2882 bp in length
* 6322 6421: gap of 100 bp
* 6422 10506: contig of 4085 bp in length
* 10507 10606: gap of 100 bp
* 10607 16241: contig of 5635 bp in length
* 16242 16341: gap of 100 bp
* 16342 19584: contig of 3243 bp in length
* 19585 19684: gap of 100 bp
* 19685 22008: contig of 2324 bp in length
* 22009 22108: gap of 100 bp
* 22109 24532: contig of 2424 bp in length
* 24533 24632: gap of 100 bp
* 24633 27409: contig of 2777 bp in length
* 27410 27509: gap of 100 bp
* 27510 31499: contig of 3990 bp in length
* 31500 31599: gap of 100 bp
* 31600 36750: contig of 5151 bp in length
* 36751 36850: gap of 100 bp
* 36851 39844: contig of 2994 bp in length
* 39845 39944: gap of 100 bp
* 39945 42665: contig of 2721 bp in length
* 42666 42765: gap of 100 bp
* 42766 45870: contig of 3105 bp in length
* 45871 45970: gap of 100 bp
* 45971 48390: contig of 2420 bp in length
* 48391 48490: gap of 100 bp
* 48491 51340: contig of 2850 bp in length
* 51341 51440: gap of 100 bp
* 51441 54723: contig of 3283 bp in length
* 54724 54823: gap of 100 bp
* 54824 56829: contig of 2006 bp in length
* 56830 56929: gap of 100 bp
* 56930 60492: contig of 3563 bp in length
* 60493 60592: gap of 100 bp
* 60593 67232: contig of 6640 bp in length
* 67233 67332: gap of 100 bp
* 67333 70877: contig of 3545 bp in length
* 70878 70977: gap of 100 bp
* 70978 73415: contig of 2438 bp in length
* 73416 73515: gap of 100 bp
* 73516 80763: contig of 7248 bp in length
* 80764 80863: gap of 100 bp
* 80864 83214: contig of 2351 bp in length
* 83215 83314: gap of 100 bp
* 83315 87723: contig of 4409 bp in length
* 87724 87823: gap of 100 bp
* 87824 93411: contig of 5588 bp in length
* 93412 93511: gap of 100 bp

```

```

* 93512 97223: contig of 3712 bp in length
* 97224 97323: gap of 100 bp
* 97324 103588: contig of 6265 bp in length
* 103589 103688: gap of 100 bp
* 103689 109134: contig of 5446 bp in length
* 109135 109234: gap of 100 bp
* 109235 112265: contig of 3031 bp in length
* 112266 112365: gap of 100 bp
* 112366 122925: contig of 10560 bp in length.
FEATURES
  source
    1. 122925
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="RP11-332019"
      /clone_lib="RPCI-11.2"
      1. 3339
        /note="assembly_fragment:00165.0"
        3440..6321
        /note="assembly_fragment:00185.0"
        6422..10506
        /note="assembly_fragment:00209.0"
        10607..16241
        /note="assembly_fragment:00218.0"
        16342..19584
        /note="assembly_fragment:00304.0"
        19685..22008
        /note="assembly_fragment:00308.0"
        22109..24532
        /note="assembly_fragment:00374.0"
        24633..27409
        /note="assembly_fragment:00451.0"
        27510..31499
        /note="assembly_fragment:00459.0"
        31600..36750
        /note="assembly_fragment:00480.0"
        36851..39844
        /note="assembly_fragment:00486.0"
        39945..42665
        /note="assembly_fragment:00514.0"
        42766..45870
        /note="assembly_fragment:00595.0"
        45971..48390
        /note="assembly_fragment:00635.0"
        48491..51340
        /note="assembly_fragment:00646.0"
        51441..54723
        /note="assembly_fragment:00658.0"
        54824..56829
        /note="assembly_fragment:00660.0"
        56930..60492
        /note="assembly_fragment:00701.0"
        60593..67232
        /note="assembly_fragment:00789.0"
        67333..70877
        /note="assembly_fragment:00838.0"
        70978..73415
        /note="assembly_fragment:00912.0"
        73516..80763
        /note="assembly_fragment:01027.0"
        80864..83214
        /note="assembly_fragment:01043.0"
        83315..87723
        /note="assembly_fragment:01079.0"
        87824..93411
        /note="assembly_fragment:01273.0"
        93512..97223
        /note="assembly_fragment:01345.0"
        97324..103588
        /note="assembly_fragment:01357.0"
        103689..109134
        /note="assembly_fragment:01384.0"
        109235..112265

```



```

* 47832 179129: contig of 131298 bp in length
* 179130 179329: gap of unknown length
* 179330 196149: contig of 16820 bp in length.
FEATURES             Location/Qualifiers
     source           1..196149
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
BASE COUNT      80057 a 19753 c 18800 g 77138 t 401 others
ORIGIN
1
20.4%; Score 79.2; DB 27; Length 196149;
Best Local Similarity 55.4%; Pred. No. 5.1e-11;
Matches 175; Conservative 0; Mismatches 138; Indels 3; Gaps 1;
QY 25 tgacacgaaatagctcttccttcgactctcttcgacttcgcaattggcttcagg 84
Db 13506 TAAAGCTATAATTGCCCTTCAGTACTTCATCCCAATATAAGCAATAGCTGAAGAA 13565
QY 85 ctgagcgatgctccacttcgagcgagattgctcgcacatgagacatcagtgaggaatt 144
Db 13566 CACAACGATGGAACTTTGGGTGACAGATGGATCCATTAGATGTTATGATATGCAAT 13625
QY 145 ttgtcccaatttaactattggcgctccagttattgaaagttgagaaagcacacaaagg 204
Db 13626 TTGTCCTAATTTATCTTTGGTCCACCTGTTATTAAATTAATTAATAAATAATACAAAA 13685
QY 205 catatttgatgacacattatggttacaaatcctcttgattatgttgaaaccttgcaa 264
Db 13686 GTATTTTGTGATGACACTTAATGGTGAATATCCAGAAAAATATGTACCATTTGTTAA 13745
QY 265 aagctggtgcttcgtgtttacatttcagtagagacataaaagataaactggaaagaaac 324
Db 13746 AAAC---ATCAATCACTAATCTTTTCATTTTGAAGCATTAATGAAGATACCGAAGAT 13802
QY 325 ttatccaaagaatcaa 340
Db 13803 GTATACAATTAGCCAA 13818

RESULT 7
SCPOS18      SCPOS18      911 bp      DNA      PLN      11-MAR-1997
DEFINITION   S.cerevisiae RPE gene.
ACCESSION    X83571
VERSION      X83571.1 GI:609673
KEYWORDS     rep gene; Ribulose-5-phosphate-epimerase.
SOURCE       baker's yeast.
ORGANISM     Saccharomyces cerevisiae
              Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
              Saccharomycetaceae; Saccharomycetes.
REFERENCE    1 (bases 1 to 911)
AUTHORS      Juhnke, H.
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 911)
AUTHORS      Juhnke, H.
TITLE        Direct Submission
JOURNAL      Submitted (19-DEC-1994) H. Juhnke, Institut fuer Mikrobiologie,
              Johann Wolfgang Goethe-Universitaet, Marie-Curie-Strasse 9,
              Gebaeude N250, AG Entian, 60439 Frankfurt, FRG
REFERENCE    3 (bases 1 to 911)
AUTHORS      Miosga, T. and Zimmermann, F.K.
TITLE        Genetic and functional analysis of the Saccharomyces cerevisiae
              gene EP11 encoding ribulose-5-phosphate 3-epimerase
JOURNAL      Yeast 11, 500-500 (1995)
REFERENCE    4 (bases 1 to 911)
AUTHORS      Miosga, T. and Zimmermann, F.K.
TITLE        Cloning and characterization of the first two genes of the
              non-oxidative part of the Saccharomyces cerevisiae
              pentose-phosphate pathway
JOURNAL      Curr. Genet. 30 (5), 404-409 (1996)
MEDLINE      97074231

* 47832 179129: contig of 131298 bp in length
* 179130 179329: gap of unknown length
* 179330 196149: contig of 16820 bp in length.
FEATURES             Location/Qualifiers
     source           1..196149
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
BASE COUNT      80057 a 19753 c 18800 g 77138 t 401 others
ORIGIN
1
20.4%; Score 79.2; DB 27; Length 196149;
Best Local Similarity 55.4%; Pred. No. 5.1e-11;
Matches 175; Conservative 0; Mismatches 138; Indels 3; Gaps 1;
QY 25 tgacacgaaatagctcttccttcgactctcttcgacttcgcaattggcttcagg 84
Db 13506 TAAAGCTATAATTGCCCTTCAGTACTTCATCCCAATATAAGCAATAGCTGAAGAA 13565
QY 85 ctgagcgatgctccacttcgagcgagattgctcgcacatgagacatcagtgaggaatt 144
Db 13566 CACAACGATGGAACTTTGGGTGACAGATGGATCCATTAGATGTTATGATATGCAAT 13625
QY 145 ttgtcccaatttaactattggcgctccagttattgaaagttgagaaagcacacaaagg 204
Db 13626 TTGTCCTAATTTATCTTTGGTCCACCTGTTATTAAATTAATTAATAAATAATACAAAA 13685
QY 205 catatttgatgacacattatggttacaaatcctcttgattatgttgaaaccttgcaa 264
Db 13686 GTATTTTGTGATGACACTTAATGGTGAATATCCAGAAAAATATGTACCATTTGTTAA 13745
QY 265 aagctggtgcttcgtgtttacatttcagtagagacataaaagataaactggaaagaaac 324
Db 13746 AAAC---ATCAATCACTAATCTTTTCATTTTGAAGCATTAATGAAGATACCGAAGAT 13802
QY 325 ttatccaaagaatcaa 340
Db 13803 GTATACAATTAGCCAA 13818

RESULT 7
SCPOS18      SCPOS18      911 bp      DNA      PLN      11-MAR-1997
DEFINITION   S.cerevisiae RPE gene.
ACCESSION    X83571
VERSION      X83571.1 GI:609673
KEYWORDS     rep gene; Ribulose-5-phosphate-epimerase.
SOURCE       baker's yeast.
ORGANISM     Saccharomyces cerevisiae
              Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
              Saccharomycetaceae; Saccharomycetes.
REFERENCE    1 (bases 1 to 911)
AUTHORS      Juhnke, H.
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 911)
AUTHORS      Juhnke, H.
TITLE        Direct Submission
JOURNAL      Submitted (19-DEC-1994) H. Juhnke, Institut fuer Mikrobiologie,
              Johann Wolfgang Goethe-Universitaet, Marie-Curie-Strasse 9,
              Gebaeude N250, AG Entian, 60439 Frankfurt, FRG
REFERENCE    3 (bases 1 to 911)
AUTHORS      Miosga, T. and Zimmermann, F.K.
TITLE        Genetic and functional analysis of the Saccharomyces cerevisiae
              gene EP11 encoding ribulose-5-phosphate 3-epimerase
JOURNAL      Yeast 11, 500-500 (1995)
REFERENCE    4 (bases 1 to 911)
AUTHORS      Miosga, T. and Zimmermann, F.K.
TITLE        Cloning and characterization of the first two genes of the
              non-oxidative part of the Saccharomyces cerevisiae
              pentose-phosphate pathway
JOURNAL      Curr. Genet. 30 (5), 404-409 (1996)
MEDLINE      97074231

```

**AUTHORS** MIPS.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org

**REFERENCE** 2 (bases 1 to 1328)  
**AUTHORS** Cziepluch, C., Kordes, E., Pujol, A. and Jauniaux, J.C.  
**TITLE** Sequencing analysis of a 40.2 kb fragment of yeast chromosome X reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2, SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes, three remnant delta elements and a Ty4 transposon  
**JOURNAL** Yeast 12 (14), 1471-1474 (1996)  
**MEDLINE** 97103775  
**REFERENCE** 3 (bases 1 to 1328)  
**AUTHORS** Miosga, T. and Zimmermann, F.K.  
**TITLE** Genetic and functional analysis of the Saccharomyces cerevisiae gene EPI1 encoding ribulose-5-phosphate 3-epimerase  
**JOURNAL** Yeast 11, 500-500 (1995)  
**MEDLINE** 97103775  
**REFERENCE** 4 (bases 1 to 1328)  
**AUTHORS** Miosga, T. and Zimmermann, F.K.  
**TITLE** Cloning and characterization of the first two genes of the non-oxidative part of the Saccharomyces cerevisiae pentose-phosphate pathway  
**JOURNAL** Curr. Genet. 30 (5), 404-409 (1996)  
**MEDLINE** 97074231

**FEATURES**

source	Location/Qualifiers
	1..1328
	/organism="Saccharomyces cerevisiae"
	/db_xref="taxon:4932"
gene	/chromosome="X"
	complement(131..847)
	/gene="POS18"
	complement(131..847)
	/gene="POS18"
	/note="ORF YJL121c"
	/codon_start=1
	/protein_id="CAA89415.1"
	/db_xref="GI:1008313"
	/translation="MWKPLIAPSILASDFANLGCCHKVINAGADWLHIDVMDGHFVP NITLGGPITVLSRVSPPREGDASNTKEKTAFFDCHMVENPEKWNVDFAKGDQFT FHYEATDGTDLVKLKSIGRAKCAIKPGTSVDVLFELAPHLONALVMTVEPFGGG KFEMDMMPVETLRAKFPHLNTQVDGGLGKETIPKAAKAGANVIVAGTSVFETAADPHD VIFMKKEVSKELSRDLLD"
	779..1102
	/note="ORF YJL120w"
	/codon_start=1
	/protein_id="CAA89416.1"
	/db_xref="GI:1008314"
	/translation="MTFATQGVSEVRKDTGSYNWFDHFLCYLALGISKWPSCKMLY RVCFILFTGGFLEWLSYNRHSRLRIKKVPIHIQKSDYFSPSTWIAVLVFSWR HIFCF"
	779..1102

**BASE COUNT** 374 a 275 c 287 g 392 t  
**ORIGIN**

**Query Match** 20.4%; Score 79; DB 45; Length 1328;  
**Best Local Similarity** 54.6%; Pred. No. 4.3e-11;  
**Matches** 231; Conservative 0; Mismatches 150; Indels 42; Gaps 2;

**Qy** 5 aaagaagaagaatgggaatgacaccgaaatagctcctctcgaatgctcctcgaact 64  
**Db** 863 AAACACACAGAAAAATGGTCAACCAATATATAGCTCCCAAGTATCTGCTTCTGACT 804  
**Qy** 65 cgcaattgctccaggctcagcgatccctcacttcggcgccgattggctccacat 124  
**Db** 803 CGCAACATGGGTGCGGAATGTCATAGGTCATCAACGCCGCCAGCTGGTTACTAT 744  
**Qy** 125 ggacatcagatgggcatgttttcccaatttaactattggcgctccagttattgaaag 184  
**Db** 743 CGATGTCATGACGCCCATTTTGTTCACCAACATTACTCTGGGCCAACCAATTTGTACCTC 684

**Qy** 185 tttag-----aaagcacacaaggc 205  
**Db** 683 CCTAGCGTCTGTGCGCCCTGGCGATGCTAGCAACACAGAAAGCCCACTGC 624  
**Qy** 206 atatttgattgacaccttattggtacaaatcctcttattgttgaaccttgcaaa 265  
**Db** 623 GTTCTCGATTGTACATGATGGTTGAAATTCCTGAAAAATGGTCGACGATTTGCTAA 564  
**Qy** 266 agctggctcttctgttttattcatttcagtagacatcaaaagataactggaagaact 325  
**Db** 563 ATGTGGTCTGACCAATTTAGTTTCCACTACGAGGCCACACAAAGACCTTTGGCATTTAGT 504  
**Qy** 326 tatccaaagaatcaatgacatgcatgattcctctggttagcattaaagcctggagcccc 385  
**Db** 503 TA---AGTTGATTAGTCTAAGGGCATCAAGCTGCGCCATCAACCTGGTACTTC 447  
**Qy** 386 cgt 388  
**Db** 446 TGT 444

**RESULT** 9  
**RCU23145** 5960 bp DNA BCT 28-OCT-1997  
**LOCUS** Rhodobacter capsulatus Calvin cycle carbon dioxide fixation operon: fructose-1,6-/sedoheptulose-1,7-bisphosphate aldolase (cbba) gene, partial cds, Form II ribulose-1,5-bisphosphate carboxylase/oxygenase (cbm) gene, complete cds, and Calvin cycle operon: pentose-5-phosphate-3-epimerase (cbbe), phosphoglycolate phosphatase (cbbz), and cbbY genes, complete cds.  
**DEFINITION** U23145  
**ACCESSION** U23145  
**VERSION** U23145.1 GI:2564972  
**KEYWORDS** Rhodobacter capsulatus.  
**SOURCE** Rhodobacter capsulatus.  
**ORGANISM** Rhodobacter capsulatus.  
**REFERENCE** 1 (bases 1 to 5960)  
**AUTHORS** Larimer, F.W., Lu, T.-Y.S. and Buley, D.M.  
**TITLE** Sequence and expression of the Form II ribulose-bisphosphate carboxylase/oxygenase (Rubisco) gene from Rhodobacter capsulatus  
**JOURNAL** FASEB J. 9, A1275 (1997)  
**REFERENCE** 2 (bases 1786 to 5960)  
**AUTHORS** Larimer, F.W.  
**TITLE** Sequence and expression of the pentose-5-phosphate 3-epimerase (cbbe) Calvin cycle operon of Rhodobacter capsulatus  
**JOURNAL** Unpublished  
**REFERENCE** 3 (bases 1 to 5960)  
**AUTHORS** Larimer, F.W.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (21-MAR-1995) Frank W. Larimer, Biology Division, Oak Ridge National Laboratory, P.O. Box 2009, Oak Ridge, TN 37831-8080, USA  
**COMMENT** On Oct 27, 1997 this sequence version replaced gi:727416.  
**FEATURES**

source	Location/Qualifiers
	1..5960
	/organism="Rhodobacter capsulatus"
	/strain="C.B. van Nieu ATH 2.3.1; ATCC catalog #11166"
	/db_xref="taxon:1061"
gene	1..345
	/gene="cbba"
	<1..345
	/gene="cbba"
	/EC_number="4.1.2.13"
	/note="aldolase; fbaB; cfbB"
	/codon_start=1
	/transl_table=11
	/function="component of pentose-phosphate cycle"
	/product="fructose-1,6-/sedoheptulose-1,7-bisphosphate aldolase"
	/protein_id="AAB82047.1"
	/db_xref="GI:727417"

```

/translation="LQDIINEFGGAMPQTFGVPVEIIVRGIKMGVRYKVNIDTCMRM
RGFRRIASONKAEDPRKFLKPAAMDAMRDLCKARLEAFGTAGHASKIKVIPMDMAK
RYASGSLAPKTN"
351. .1740
/gene="cbbm"
RBS
351. .354
/gene="cbbm"
364. .1740
/gene="cbbm"
/EC_number="4.1.1.39"
/notes="rbpL"
/codon_start=1
/transl_table=11
/function="carbon dioxide fixation"
/product="Form II ribulose-1,5-bisphosphate
carboxylase/oxygenase"
/protein_id="AAB82048.1"
/db_xref="GI:727418"
/translation="MDQSNRYARLDLKEADLIAGRHVLCAYIMPKAGYGYLETAH
FAEASTGTNVSTDDTFRGVDAIYVIDPEKEIMKIAYPVLEFDRNIIDGGAMLC
SETLTIIGNQMGDVEYAKMHDFVPPCYLRLEDFGSPMNIADMRVLRPVVDGMV
VTTIKLGLRKPFDACYEFWLGDFIKNDPQGNOTFAPLKETIRIVADAMKRA
ODETEALFSANITADDHYEMVARGEVILETFGENADHVAFLVDGYVTGPAITPAR
RSFPQFLHYHRAGHCYAVTSPQSMRGYTAFLVSKSRLOQASGIHTGTMGYKMEGDA
SDKIMAYMLNDEAAGPPYHODWLOMKATPTIISGMNALRLPGFDNLGHSNVIQTS
GGGAFGLDGTAGAKSLRQSCDAWKAGVDIVTYAKSHRELARAFESFPNDADKLYPG
MRVALGVN"
1755. .1776
/rpt_type=inverted
repeat_unit
1786. .2483
/gene="cbbg"
RBS
1786. .1790
/gene="cbbg"
1797. .2483
/gene="cbbg"
/EC_number="5.1.3.1"
/notes="ribulose-phosphate epimerase; cfxE"
/codon_start=1
/transl_table=11
/product="pentose-5-phosphate-3-epimerase"
/protein_id="AAB82049.1"
/db_xref="GI:2564973"
/translation="MTFDRSIKTAPSILSADFADFGRFETQAIQAQAQADWNVHVDG
FVPLNFTGPPAVAFKRVHTVMDVHLMISFVDAYIDAYAQADVLTAHVEAGPHI
RLQAIKRAKSGVAINPGTPAEIEHVDIADVVCVMTVNFPGQYQYIDMTAKVR
KLRMIGDRPVHIEIDGMDPVTAPLMAAAGADVFAVAGSVFKGSGSVTPVYVGNIA
AIRAAQAAL"
2492. .3043
/notes="hypothetical protein; similar to A. brasiliense f1cA
gene, encoded by GenBank Accession Number Y12363,
Synecocystis sp. ORP_ID:s111289, encoded by GenBank
Accession Number D90904, M. tuberculosis MTCY04C12.12 orf,
encoded by GenBank Accession Number Z81360, and M. leprae
hypothetical protein MLCB2052.40, encoded by GenBank
Accession Number Z98604"
/codon_start=1
/transl_table=11
/protein_id="AAB82050.1"
/db_xref="GI:2564974"
/translation="MAVSPVPCDFGNKAPDFRLPDFSGRFNSLIDIRGPKGLLIIMFIC
NHCPYVQAILDRQDAEALMDLIGVAAISSNDVASYEDFSFANMARLAABRGFRFP
YLDETQAVAKAYDAICTPDPFGFNAECQLQYNGRLDASGRTPGPADARRELYEAMLS
VAETGQGPQRIASIMGCSIKWKE"
3043. .3702
/gene="cbbz"
3043. .3702
/gene="cbbz"
/EC_number="3.1.3.18"
/codon_start=1
/transl_table=11
/product="phosphoglycolate phosphatase"
/protein_id="AAB82051.1"
/db_xref="GI:2564975"

```

```

/translation="MAPALIFDLDTGLSDSAPAIHKVSNVDVLRAGYAPLGLDQIRSF
VQGAHPLVRCLLTTAGDEPEGLFDAYADLVSRYETDVEGNTLYPGVITALQRLRE
MGCPLAITNPKYKPAALAHVGLTDFQVLVIGDGLTPKPNPMEVNEARRVLRP
HALYIGDSEIDAQTAQNAQLPFIYTEGYRKTPLDALPHAAKFHDFSALPGIVEGWTW
S"
RBS
3682. .3685
/gene="cbbY"
3682. .4376
/gene="cbbY"
3693. .4376
/gene="cbbY"
/codon_start=1
/transl_table=11
/product="CbbY"
/protein_id="AAB82052.1"
/db_xref="GI:2564976"
/translation="MELKALIFSDVDGTLAETEEVHROAFNETFAAOGLDHWWSKEDYR
TLIRTTGCKERMAKHRENILSGSPSDAKIADLHKAKTQRYVEIITASGOVGLLPVGAELI
DRAKASGLRLAIAITTTTANYDALIAITFSKPADGIDFEVIAAGDEVAAQKPAQPVYLR
ALQGLGLPAACLAPEFDSRAGLSARAAGLRVLTPTSEYTRGDDFSADWRIPDLASA
ATQAPELP"
4437. .4877
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="AAB82053.1"
/db_xref="GI:2564977"
/translation="MDLRTLPDFAVAPQIALDIAEAARLGFRTLINRNPDEVESG
LNAALQAAAAAGLSVYHLPPYPGELTPDLIAGFEAALAAADAPVLAWCRRSGTSSHL
WALSOQORPLEEIVQTAQNAGYDLSALVPMRLAKAADSSGVIV"
5054. .5806
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="AAB82054.1"
/db_xref="GI:2564978"
/translation="MLDHLKLLVADDIATINSHCAQMAKFGDLRIIVGTFRFSATQ
NLGSLDPTLLISNHDPAFLNSFLGAGLFSEAPMVKWAIDSVGACSWRMAERNAAGEL
TEGERKVIELNRRFQVLGYSIGFPDTSARAKAIGLCAQPGLTQADVDAIWDEHGRD
LRVLSIMHLKISALPSTFTRRLTARQREVLEWVGDKRTSADIAQLGLTTATVEKH
LRLAREALEVETTAQALLKASKMQLFITSAP"
BASE COUNT 1123 a 1908 c 1857 g 1072 t
ORIGIN
Query Match 20.3%; Score 78.6; DB 2; Length 5960;
Best local Similarity 55.8%; Pred. No. 6e-11;
Matches 150; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
QY 33 aaaaatagctctctcgatctcttcgcagattcgccaatcttcgaggtcagcgc 92
|| || || || || || || || || || || || || || || || || || || ||
Db 1818 AAGATCGCCCTCTGATCCTTCGCGGACCTCGCCGCTCGGCGCTGAAATCCAGGC 1777
QY 93 atgtccattcggcgccgattggtctccacatggacatcatgcatggtgcatgttcccc 152
|| || || || || || || || || || || || || || || || || || || ||
Db 1878 ATCGAGGCGCAGGCGCGATTGGTGCATCTCGACCTGATGATGGGCATTCGTGCGCG 1937
QY 153 aatttaactattgcygtccagttattgaaagtttgaaagacacacaaaggcatatttg 212
|| || || || || || || || || || || || || || || || || || || ||
Db 1938 AACCTCACCTTTGTGCGCGCGCGCTCGCGGCTTCGCGCAAGCATGTGAAAACCGTGATG 1997
QY 213 gatigtacaccttgggttacaataacctctctgattattgaaaccttggcaaaagctggt 272
|| || || || || || || || || || || || || || || || || || || ||
Db 1998 GACGTGCATCTCATGATCTGCGCGGTGCATATATATGACGCTTATGCGAGCGGCG 2057
QY 273 gctctgtgttttacatttcacgtagagac 301
|| || || || || || || || || || || || || || || || || || || ||
Db 2058 GCCGATGTGCTGACCGGCCCATGTGCGAGGC 2086

```

```

RESULT 10
AE002306
LOCUS AE002306 12939 bp DNA BCT 26-MAY-2000

```

Chlamydia muridarum, section 37 of 85 of the complete genome.  
AE002306 AE002160  
VERSION AE002306.2 GI:8163215  
SOURCE Chlamydia muridarum.  
ORGANISM Chlamydia muridarum  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
1 (bases 1 to 12939)  
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Genome Sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
20150255  
2 (bases 1 to 12939)  
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Direct Submission  
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
On Jun 1, 2000 this sequence version replaced gi:7190428.  
Location/Qualifiers  
1. .12939  
/organism="Chlamydia muridarum"  
/db\_xref="taxon:83560"  
/note="synonym: Chlamydia trachomatis MoPn"  
complement(115..1941)  
/gene="TC0388"  
complement(115..1941)  
/gene="TC0388"  
/note="similar to GB:232522 PID:510140 SP:P54124  
PID:1771208; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="oligodeoxytase F"  
/protein\_id="AAF39245.1"  
/db\_xref="GI:7190429"  
/translation="MTTDTTSQTLVRSKRDVPSDCWDTKSLYASRAWKKDLNLCG  
KEAPFPHNEQFHEOPSLYKLLTEVEFISLEKLYVYVAHLTHDDISNQATA  
DLKSYLLTSFSEELSWIOPALJALPQKISELLTSPLOYHFVLYKIPRLAPHTG  
TSREKILASSPALVAYKTFSSLDSDIPIGGEAVDSGSKSHPLSHLSLYMQSPD  
RELKSAQKQCYRHYGLSLANLNGIKQALHFNKAKHNTKSLERAAALFONNISTS  
VVTTLTYKHHTLTKYFQKQKALGDPDFHYDVIAPAASEFARHYSEEAVAL  
ICDLPPLAGDNYEALRKLGTSEGWDYENQNKRSYSSGCDYKPYLLNYTGL  
YDVSVVAHEGSHMSHFLSHKHQYHEAQYPIFLAEIATNLMEPLKLNAPSKE  
EKTAILSRSLDTIFATLFRQTLFADPELAHSAEQGIPLEEFFSQSYKKLOHIFYG  
DVTTFELSIENARIPHEFYNYVYATGIIASCFSEKILSKAQAQAYLTFLO  
SGSDPFIEILKSGLDMTSSAPMLAAFSYIERKLDLTNLL"  
2237..4840  
/gene="TC0389"  
/note="similar to GB:M29364 SP:P03815 GB:V00350 GB:X57620  
PID:1236633; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent Clp protease, subunit B"  
/protein\_id="AAF39246.1"  
/db\_xref="GI:7190430"  
/translation="MEKFSADVSEALEKAPLAKDAKHSVVTENHLLKSLQNPGLSF  
CLVKKVHGNLGLTSDALHREPTVVEGAILPKPSLQLOLLLNQAEARSMDGE  
YLSGDHLLALFATKEPFAWRKTKTSPALKELIILKQSRMDSFAENLKL  
EKYCKNTILAREKGLDPVIGRDEIRRTIQVLSRRKNNMPLIGPGVGRKTAIAEGL  
ALRTVQGDVPESLKHLVYLDGALIAAGYKRFEEERKLSVKLKGVEASEGACILFI  
DEVTPLGACATGDGMAANLLKPALARGTLHCIGATTLNEYOKYIEKDAALERRFPQ  
IFVTEPSLEDAVFLRLGKREYIEFHGVRITEGALNAAVLSYRITDRFLPDKAIDL  
IDEASLIRMQISGLPLPIDEKERLSALIVKQEAIKREQAPAYQAEANEMQKADRV

KEELAALRLRWDEEKGILAGLKEKKNSLENLAFEEAEARTADYNKVAELRSLIPSL  
EKIRLAEEALNORDGRLLQEEVDERLIAQVVAWNTGIPVQKMLEGESKLLVLESFL  
EERVQGPFAITAVSDSISARVGLSDPQPLGVFLGPTGVGTAKALAEALLFN  
KEAMINFDMEYMEKHSVKLLGSPGPGVYEGGSLSEALRRRPSYVVLDELEKA  
DKEVFNILQIFDGLTDSKKRKNCKNALFIMTNSIQSELALCAKKGITVDEKA  
VLSVAPALKNFYSEFINRDIPLFVLTETDVIKVIQIMNRVALRLLERRISLT  
WDDSVVFLFSEQYDGAFGARPLKRLIQKVVVMTLSKALLKGDIKSGMSVELTMAKDV  
VFKTKTNSVV"  
4850..6325  
/gene="TC0390"  
4850..6325  
/note="TC0390"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF39247.1"  
/db\_xref="GI:7190431"  
/translation="MMKHWRIRIIGISLSPVLLALGEXTPAFFLERNRNFWNIDPY  
CLISACFTSNODPLSAERLVLFPLQSOEDLLVFAKYLSSKNAYLFSOAEERIF  
AKTALPRVSLGNCNRADDLAKVLAEGASTAKQKARYSLYLDVALRAYVBRARLV  
NATCEVASQIDIASIEAINTLFEKSGVRYPSEKMFESRSELAATVDSKFGVCLGS  
VVLQVAQRNLNLSLEAVTPPGHIYLRKDVVNIETSSGRHIPTKYECEIPESOLR  
VROEELIGLTFMNRGAFFLQKGEFQAISYAEQAKYLADQISDLGITYILLGKK  
KEGEDILKSSERTKRKSAIHDFERNISPEILAVLFADSGVTYQETLDYHKLLVIV  
EKPRSGALRLAATALELGLVKGEQLLEKSEVEDAPDDLFLRLQLCKILCNRHDT  
RAKHFDQAKKILAEEGMLSESSYTLKLEKKLALVSPH"  
6538..6912  
/gene="TC0391"  
6538..6912  
/note="TC0391"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF39248.1"  
/db\_xref="GI:7190432"  
/translation="MTKICSTSQGLAACRIERINLEKSLNPRNAAPTTRKLTJVALA  
AISAITFLVIAGLLFVGVLCPPVLIASALLGLGAFFLGTGTVVGLLTVEAAAKDRL  
YRSQVLSNNLCYKTAGVIRNG"  
7009..7440  
/gene="TC0392"  
7009..7440  
/note="TC0392"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF39249.1"  
/db\_xref="GI:7190433"  
/translation="MEGIRLFCCKDHLCKLVNPNVCSLKKKTCERKQVQVAGCIGVI  
CSMVCLALGIAAGTGLLGAFTVSLGVALLGLVLFMSIYDVLERHGGILGCPKMT  
REPAPEPAPEPAPELSVKTGKGQEDCAVIYGCCK"  
7452..7736  
/gene="TC0393"  
7452..7736  
/note="TC0393"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF39250.1"  
/db\_xref="GI:7190434"  
/translation="MIQSTIEGTFIESHRRPCDLSKHKAKAAGFIKLIATLVALIL  
NGALPALSIVIAVCGVSTPASPLVGLAATLASFLCAARVFLITKDRGW"  
7771..8271  
/gene="TC0394"  
7771..8271  
/note="TC0394"  
/gene="TC0394"

```

/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39251.1"
/db_xref="GI:7190435"
/translation="MDGRMVCCGKILLGEQOTLPTVERSPSSICATIVVKTKALLAVSL
FAVLTAGSLVLSGLVFSGTASFPLPYLLIFAAALFGVCASIVLVKCISSVVQDVKK
KEKAPESIPSDLSLSSSSSELEKLVDSSSLGSLISIGDCVDEAFGRNRELARMR
ARANS"
8207..8383
/gene="TC0395"
CDS
8207..8383
/gene="TC0395"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39252.1"
/db_xref="GI:7190436"
/translation="MKHLGTMSCGLGLEQIAVNNRRGGEPFLFYHFRSLICPCN
KVLISQGFFFKVA"
8207..8383
/gene="TC0396"
CDS
8207..8383
/gene="TC0396"
/note="similar to GP:3172545; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="inclusion membrane localised protein Inca"
/protein_id="AAF39253.1"
/db_xref="GI:7190437"
/translation="WNSPLVEMPLSCYPTHTSACTKRSSLYKPSLIETVQVRA
AFVSLAISLIGLAIGHAIGLAPQALVLIATFIISLGNALYICKAPLRKLV
AQEQVASKVEFLNFKSKEQFGLSKDFATSKDLSVLDHNLQDFQSSHQGF
EDLEEDKNSAELRQIFQSVQSLKSTILSKBEIKIIVPLTEVFRLEAKENEDLL
KIVQDLQIRDKLRABINNLQSQASTLSQIASQIEENEKLYANITKALSQSYSDSSA
"
9703..10389
/gene="TC0397"
CDS
9703..10389
/gene="TC0397"
/note="similar to PID:606320 PID:1789788; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribulose-phosphate 3-epimerase"
/protein_id="AAF39254.1"
/db_xref="GI:7190438"
/translation="MKKQGVLIAPISIMGADLACLGDAAARNIEESGANLIHIDVMDGHF
"
Query Match 20.0%; Score 77.6; DB 1; Length 12939;
Best Local Similarity 53.4%; Pred. No. 1.2e-10;
Matches 186; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

Qy 35 aatagctcttcgatgctcttcctccgaactcgccaattggcttcgcagagctcagcgcat 94
Db 9723 AATTGCTCCATCATATTATGGGGCCGACATYTGCTTCCTAGGGATGACGACGAAATAT 9782

Qy 95 gctccacttgccgcagctgctccacatggacatcatggatggcctatttggctcccaa 154
Db 9783 AGAAGAGTCGCGGCAAACTTATTATATAGATGTCATGGATGGCATTTTGTTCGAA 9842

Qy 155 tttaacctattggcctccagctatgaagtgtgagaagacacacacagagcatattggga 214
Db 9843 TATTACCTTGGTCCGGTATTATGCCCAATCAATCGGTCAACAGATCATTTCTCTTGA 9902

Qy 215 ttgtccacctattgttacaaactctcttattgttgaaccttggcaaaagctggtgc 274
Db 9903 AGTTTCATGCTATGATTATACGCCGCTTGTGAATTTGTAGAGGCTTTTGTAAAGCTGGAGC 9962

```

```

QY 275 ttctggttttcatcttcacgtagacacatcaaaagataactggaagaacttatccaaag 334
Db 9963 AGATCGTATTATTTGACCTTTGAGGCAGC---AGAAACCTTTAAAGAAATCTTTGATTA 10019

QY 335 aatcaagtcacatggcgcattcctggtgtagcattaaagcctgggac 382
Db 10020 TATTCGAAATGTGGAGTCAGGCAGGAGGATCGCTTTTCTCCAGAGAC 10067

RESULT 11
AE002217 10150 bp DNA BCT 30-MAY-2000
LOCUS Chlamydomophila pneumoniae AR39, section 45 of 94 of the complete
DEFINITION genome.
ACCESSION AE002217 AE002161
VERSION AE002217.2 GI:8163462
KEYWORDS
SOURCE Chlamydomophila pneumoniae AR39.
ORGANISM Chlamydomophila pneumoniae AR39
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 10150)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 2 (bases 1 to 10150)
REFERENCE Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
MEDLINE Medical Center Dr, Rockville, MD 20850, USA
REFERENCE On Jun 1, 2000 this sequence version replaced gi:7189493.
FEATURES
Location/Qualifiers
1..10150
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(122..1294)
/gene="CP0581"
complement(122..1294)
/gene="CP0581"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38399.1"
/db_xref="GI:7189494"
/translation="MSSPVNTPSAPNIPAPTPPGITTPKRSSEFIEKVIIVAKYI
LFAIATGALGTILGSLGALTGGIATLIVFFVSMVLLGLLKLKDSISGSEERRLRE
EVSRTSENORLTVITTTLETKDKAKDOLLEIEAFRNENGNLKTAEDELEVQ
SKLSPOLEALERINQLOIANAGDAQEISSLKLIQSWDSKVVEQINTSQALKVLGG
QEWQEAOTHYKMQEQIQAOLAEILGHNSHNSALQSVENLVQDALFRVYVGLLE
SENKUSQACSLAROEIKLAQHETSLQORIDAMLAQEQNLAEQVTALEKMKQAKAE
SEFTIACVDRDTFGRRTPPTPTTPVVEGDEGEGTTPPVSPSPVDRATGDGQ"
complement(1349..1456)
/gene="CP0582"
complement(1349..1456)
/gene="CP0582"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11

```

gene	/product="hypothetical protein" /protein_id="AAF38400.1" /db_xref="GI:7189495" /translation="MLTSVTFIKHQISEIKMIKYLIDLPQERYKFLDF" 1742..2431 /gene="CP0583" 1742..2431 /gene="CP0583" /note="similar to PID:606320 PID:1789788; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="ribulose-phosphate 3-epimerase" /protein_id="AAF38401.1" /db_xref="GI:7189496" /translation="MKKQESVLVGPISMGADTLCLGVEAKKLEQAGSDFIHIDIMDGH FVFNLFQGIITAAINRSDFLFEVHAMINPFEIESFVRSAGDRIIVHFEASEDIK ELSYIKKGVQAGLAFSPDSTEFPLPFCDDVVLMSVPGFTGQSFPLPNTIEKI AFARHAIKTLGLKDSCLIEVDGGIDQOSAPLCRDAGADILVTASYLFEADSLAMEDKI LLRGENVGVK" 2418..2975 /gene="CP0584" 2418..2975 /gene="CP0584" /note="similar to PID:1000357 SP:P49778 PID:1303902 GB:AL009126; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="translation elongation factor P" /protein_id="AAF38402.1" /db_xref="GI:7189497" /translation="MWLSQSLQSGMFISTKDGLYKVTSVKAGPKGESFIKVALQAA DSDVTERNEKATQEVKEAGQFERTLEYLVLEDESYLEFDLGNVEKLFIPQETMKDNF LFLKAGTVSAMVDNVFVSELPHELELMVSKTDPGDSLSLGGVKRALLETGIEV MVPFVEIGDVIKIDRTCEYIQRV" 2998..3501 /gene="CP0585" 2998..3501 /gene="CP0585" /note="similar to GP:1055245; identified by sequence similarity; putative" /transl_table=11 /product="acetyl-coenzyme A carboxylase, biotin carboxyl carrier protein" /protein_id="AAF38403.1" /db_xref="GI:7189498" /translation="MDLKQLEKLIAMGRNCKFAIKREGLELELERDPRGNROEP VFYDSRLFSFGSPRIPTDPKDDTIKETTTENSESTSSGDFISSPLVGVFYQSP APDPSFVKPGDIVSEDTIIVCAEMKVMNEVKAGMSGRVLEIITNGDPPVQFGSKLF RIAKDAS" 3498..4862 /gene="CP0586" 3498..4862 /gene="CP0586" /note="similar to PID:1055246 SP:P49787 GB:AL009126; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="acetyl-coenzyme A carboxylase, biotin carboxylase" /protein_id="AAF38404.1" /db_xref="GI:7189499" /translation="MKKVLIANRGEIAVRIIRACHDGLGLSTVAVYSIADQALHVLIA DEATICEPQAASKYLKISNIIAAECITGADAVHPGYGFLSEANFASICESGGLTPI GPSSEIATMGDKTAAKSLAKIKCPVIGPSEGIIEDESEGLKTAERKIGFPIVKA GVSGRGIRVKEKDEFVFAAARAEEAGFNPNVIEKFIENRPLEHLEQVIGDTHG NVYHLGRDCTIORRROKLIETSPSILNAEIRVKVGKIAVDLARSAGYFSGVFEFL LDKDKYFFEMENRTRIOVEHTITEEVTCIDIVKEOIHVAMGNKLPWKOKNLEFSCHII QCRNAEDPNNKSPSGRLDYILPPAGPSIRVDGACYSYGAIPPYDYSIMAKVIAG KNREEATAIMKRALKEFHIGGVQSTPIPHQFMLDNPRFLSNYDINIIDLNAQGNFSFKEF" 5018..5419	gene
CDS		gene
gene	/product="hypothetical protein" /protein_id="AAF38400.1" /db_xref="GI:7189495" /translation="MLTSVTFIKHQISEIKMIKYLIDLPQERYKFLDF" 1742..2431 /gene="CP0583" 1742..2431 /gene="CP0583" /note="similar to PID:606320 PID:1789788; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="ribulose-phosphate 3-epimerase" /protein_id="AAF38401.1" /db_xref="GI:7189496" /translation="MKKQESVLVGPISMGADTLCLGVEAKKLEQAGSDFIHIDIMDGH FVFNLFQGIITAAINRSDFLFEVHAMINPFEIESFVRSAGDRIIVHFEASEDIK ELSYIKKGVQAGLAFSPDSTEFPLPFCDDVVLMSVPGFTGQSFPLPNTIEKI AFARHAIKTLGLKDSCLIEVDGGIDQOSAPLCRDAGADILVTASYLFEADSLAMEDKI LLRGENVGVK" 2418..2975 /gene="CP0584" 2418..2975 /gene="CP0584" /note="similar to PID:1000357 SP:P49778 PID:1303902 GB:AL009126; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="translation elongation factor P" /protein_id="AAF38402.1" /db_xref="GI:7189497" /translation="MWLSQSLQSGMFISTKDGLYKVTSVKAGPKGESFIKVALQAA DSDVTERNEKATQEVKEAGQFERTLEYLVLEDESYLEFDLGNVEKLFIPQETMKDNF LFLKAGTVSAMVDNVFVSELPHELELMVSKTDPGDSLSLGGVKRALLETGIEV MVPFVEIGDVIKIDRTCEYIQRV" 2998..3501 /gene="CP0585" 2998..3501 /gene="CP0585" /note="similar to GP:1055245; identified by sequence similarity; putative" /transl_table=11 /product="acetyl-coenzyme A carboxylase, biotin carboxyl carrier protein" /protein_id="AAF38403.1" /db_xref="GI:7189498" /translation="MDLKQLEKLIAMGRNCKFAIKREGLELELERDPRGNROEP VFYDSRLFSFGSPRIPTDPKDDTIKETTTENSESTSSGDFISSPLVGVFYQSP APDPSFVKPGDIVSEDTIIVCAEMKVMNEVKAGMSGRVLEIITNGDPPVQFGSKLF RIAKDAS" 3498..4862 /gene="CP0586" 3498..4862 /gene="CP0586" /note="similar to PID:1055246 SP:P49787 GB:AL009126; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="acetyl-coenzyme A carboxylase, biotin carboxylase" /protein_id="AAF38404.1" /db_xref="GI:7189499" /translation="MKKVLIANRGEIAVRIIRACHDGLGLSTVAVYSIADQALHVLIA DEATICEPQAASKYLKISNIIAAECITGADAVHPGYGFLSEANFASICESGGLTPI GPSSEIATMGDKTAAKSLAKIKCPVIGPSEGIIEDESEGLKTAERKIGFPIVKA GVSGRGIRVKEKDEFVFAAARAEEAGFNPNVIEKFIENRPLEHLEQVIGDTHG NVYHLGRDCTIORRROKLIETSPSILNAEIRVKVGKIAVDLARSAGYFSGVFEFL LDKDKYFFEMENRTRIOVEHTITEEVTCIDIVKEOIHVAMGNKLPWKOKNLEFSCHII QCRNAEDPNNKSPSGRLDYILPPAGPSIRVDGACYSYGAIPPYDYSIMAKVIAG KNREEATAIMKRALKEFHIGGVQSTPIPHQFMLDNPRFLSNYDINIIDLNAQGNFSFKEF" 5018..5419	CDS
gene	/product="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38405.1" /db_xref="GI:7189500" /translation="MHELFKIDGVYFFKFKMLFYNNYSLNSHHHEKPSLEKAVOAL DSYFYWGDDTDLARDDISREIYCVRRLYIRFWIVISQSLSRIPWRLKRILLRYCT LRGKYVMPILIKRIAIALLGLIRFSRLKSVY" 5287..5841 /gene="CP0588" 5287..5841 /gene="CP0588" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38406.1" /db_xref="GI:7189501" /translation="MASETPYQIILHAQREVDRDAYFNQADCHPARANQILEAKKICLL DVYTHNYGVFTPCVDNYPNLRFTFVSSKNMNGSLNPLDNVLVAMVRRTHARLL ACKIRINIVPRVGLDLRSGILISKLELKQPOFQSLTDFEVNHSTQBEARVHQKHV LLILLLILCKQAXLESFQEKRSS" 5790..6140 /gene="CP0589" 5790..6140 /gene="CP0589" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38407.1" /db_xref="GI:7189502" /translation="MOAGRSGIIPGKKAILNVNDAKTPNYSCIFESIGFFNPDLEA OHNQAAALVRKILKVVPHHFLKGLIAKLPSLKKDRKFMSSLIFTKLSYALDLSAPMH LEGPNLSVEEKLD" 6141..6407 /gene="CP0590" 6141..6407 /gene="CP0590" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38408.1" /db_xref="GI:7189503" /translation="MKEYLDPLVQRNVVERDPQTKRHCTVSQKFGGESIDAKTTTGQLF HIAGKTEPGHGKLCGLGESILKQLLALGIITGYENRERENVVYLD" 6587..6679 /gene="CP0591" 6587..6679 /gene="CP0591" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38409.1" /db_xref="GI:7189504"	gene
CDS		CDS
gene	/product="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38409.1" /db_xref="GI:7189504"	gene
CDS		CDS
gene	/product="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38409.1" /db_xref="GI:7189504"	gene
CDS		CDS
gene	/product="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38409.1" /db_xref="GI:7189504"	gene
CDS		CDS
gene	/product="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38409.1" /db_xref="GI:7189504"	gene
CDS		CDS
gene	/product="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF38409.1" /db_xref="GI:7189504"</	





```
gene
CDS
    complement(6874..7728)
    /gene="CpN0177"
    complement(6874..7728)
    /gene="CpN0177"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18330.1"
    /db_xref="GI:4376446"
    /translation="MKOPMSLIFSSVCLGILGSLSCNOKPSSWVNHNTSTSEPFVH
    GNKVSQHPHPSAFRTQIFSEHNDPVYVAKTDESRKIWEIHKNLKIKGSYIPI
    STYSLMHPKSAALTITKTRPHIPWINGYERSNIDTKYKNGSRRTSHDQPNRA
    VNLTKSSGRCAIGLMEETDEEDVIARRREGVTSLPVEVCSIPGNGFVYAYAWIA
    DESACSKVLVPKGYSLVWESVSSDSLNAFDSFAEDYLRTFLANGTSILCVHES
    YKKVPPQP"
    complement(8055..8321)
    /gene="CpN0178"
    complement(8055..8321)
    /gene="CpN0178"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18331.1"
    /db_xref="GI:4376447"
    /translation="MKEYLDFLVORNVVERDPQTKRHCTVSQKFGGESIDAKTTTGOLF
    HAGTEPHGKLCIGLESILKLLALGIITGYENREREWVYLD"
    complement(8322..8672)
    /gene="CpN0179"
    complement(8322..8672)
    /gene="CpN0179"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18332.1"
    /db_xref="GI:4376448"
    /translation="MOAGRSGLIIPGKAILLVNDAKTPNYSCIFESIGFNFEOOLEA
    QHNQAAVLVRKILKVPHHFLGLIAKLPRSLKKRPMSSLIETKLSYALDSAPMH
    LECKPNLSYEKLD"
    complement(8621..9175)
    /gene="CpN0180"
    complement(8621..9175)
    /gene="CpN0180"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18333.1"
    /db_xref="GI:4376449"
    /translation="MASETYPSQILHAQREVVDVFNQADCHPARANOILEAKKICLL
    DVYHTNHSYFTFCVDNYPNLRFTFVSSKNEMGLSNPLDNVLVEAMVRKTHARNLL
    AACKIRNIEVPRVGVGLDRSGILSKLELKQPFQSLTDEDFVNSTQNEEARVHQKHV
    LLISLILCKQAVLESFOEKKRSS"
    complement(9043..9444)
    /gene="CpN0181"
    complement(9043..9444)
    /gene="CpN0181"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18334.1"
    /db_xref="GI:4376450"
    /translation="MHLEFKIDGVYFFKFKMLFYNNYSLNSHHEKPSLSLEKAVQAL
    DSYFYWGDDTDLVLRADTSREIYCVRRRLYIRFVIVSISQSLSRIPFWRLKRILLRYCT
    LRGYVMPILIKRIALLGLIRFSRLKRSYV"
    complement(9600..10964)
    /gene="accC"
    /note="CpN0182"
    complement(9600..10964)
    /gene="accC"
    /codon_start=1
    /transl_table=11
    /product="Biotin Carboxylase"
    /protein_id="AAD18335.1"
    /db_xref="GI:4376451"

/translation="MKKVLIANRGEIAVRIRACHDLGLSTVAVYSLADQALHVLVA
DEAICIGEPQAAKSYLKINILAAACEITGADAVHPGYGFLSENANFASICESGLTFFI

Query Match      19.7%  Score 76.4;  DB 1;  Length 12973;
Best Local Similarity 54.7%  Pred. No. 2.5e-10;
Matches 152;  Conservative 0;  Mismatches 126;  Indels 0;  Gaps 0;

QY 35 aatagctcttcgatgctcttcgcgacttcgccaatttgcttcgagctcagcgcat 94
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12697 AGTTGCCCTTCGATATGCGGGGAGATCTTACCTGTTGGGTGTAGAACACAAAAAACT 12638
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 95 gctccacttcgcgccgattggtccacatggacatcatgcatggtggcattttgtcccca 154
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12637 AGAGCAGCGGGGAGTATTTTATACATAGATATCATGATGCCACATTTGTCGAA 12578
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 tttaactattggcgctccagttattgaaagtttgagaagcacacaagcattattgga 214
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12577 CCTTACTTTTGTGTCAGGGATCATTTGCCATTAATAGATCTACGGATCTATTTTGA 12518
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 215 ttgtcacttatggttacaaatcctcttgattgattgtaacccttggaagcctggcagcgtgc 274
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12517 AGTCCACGCTATGATTACAAATCCTTTTGAATTTATAGAAAGTTTCGTTCTGTCG 12458
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 275 ttctggtttacatttcacatccatgtagacatcaaaagata 312
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12457 GGATAGATATAGTACACTTTTGAAGCTTCAGAGGATA 12420
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
AP002545/c 300050 bp DNA BCT 07-JUL-2000
Chlamydomophila pneumoniae genomic DNA, complete sequence, section 1/4.
ACCESSION AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
AB036074 AB036075 AB036076 AB036077 AB036078 AB036345 AB036346
AB038347 BA000008
AP002545.1 GI:8978373
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
    /organism="Chlamydomophila pneumoniae"
    /strain="J138"
    /db_xref="taxon:83558"
    /note="synonym:Chlamydia pneumoniae"
    complement(1..282)
    /gene="CPJ0001"
    complement(1..282)
    /gene="CPJ0001"
    /codon_start=1
    /transl_table=11
    /product="Ct001 hypothetical protein"

gene
CDS
```

```
/protein_id="BAA98211.1"
/db_xref="GI:8978374"
/translation="MLGKIIRGLSLVILCALNVGLIGITHNKNIITAKLCGGVSTP
AQQYIIIGIAGVICLLSFCPCSKSRHSHGSDSCSGGCHSHSDKN"
573..878
/gene="gatc"
CDS
573..878
/gene="gatc"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (C subunit)"
/protein_id="BAA98212.1"
/db_xref="GI:8978375"
/translation="MEQHLDRREITLLAKASALQSLSEELIQEYQTSLSAVITSKKEA
LAIEDDADSCSLFWHVNVNEDLREDSVTSDFNREEFLRNVPESLGLVAVPAVK"
895..2373
/gene="gatA"
CDS
895..2373
/gene="gatA"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (A subunit)"
/protein_id="BAA98213.1"
/db_xref="GI:8978376"
/translation="MYRYSALELAKAVTLGELTATGVTOHFFHRIEAEQGVAFISL
AQVLEAEQLELDKKRSRGEPLGKLAGVPVGIKONIHVTKLTTCASRVLENYQPPD
ATVEERIKKEGILLKLNMBEFANGSTLISAPHTNPWDLRSKVPGGSGGSAAV
SARFCVALGSDTSGSIRQPAFCGVGVGPKPSYGAVSRYGLVAFASLDQIGPLANTV
EVALMMDVFGSDPKDATREFFRDSFMSKLSIEVPKVIQVPTFLEGLRDDIRENF
FSLAIFEGESTHVDVDELILSHAVSIYIILASAAATNLARFDGVRYGVRSPQAH
ISQLDYLRGEGFGEVMRRLILGNVLNVAERONVYKKATRAKIVKAPRTAFEC
ETLAMPVSSPAFETIGETILDVPTVLODIYVANNVLAIPALVPAVSPGSKGLPLGQ
LIQOGQDQVCQVCYSQEAHQIKQLFSKRYAKSVVLGGQS"
2370..3836
/gene="gatB"
CDS
2370..3836
/gene="gatB"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (B subunit)"
/protein_id="BAA98214.1"
/db_xref="GI:8978377"
/translation="MSAVYADWESVIGLEVHVELNTASKLFSSALNRFGEPPNTNIST
VCTGLPGLSVLQNSAVEKAVLECAVEGEISLRSRDRKSYFPDPSRNFQITQFHE
PIARGGRKAIYQGEERYFELAQTHIEDDAGMLKHFGFAGVDYNRAGVPLIEIVSKP
CMFCPEDAVATSLVSLDDYIGISDCNMEGSRFDVNVSRPKSPPELRNKVEIKN
MNSFAFMAQALEAKQRQIDYELNQPNKPKLVIPTAATYRWDPEKKTKTVMRLKESAE
DYKYPPEPLTLQLTESIYIERIKTLPELPDYKHRYIQEYGLSEDIATLISDKNI
ATFFEVACKDNFRSLNWSVVEFGGRCKTLGKLPSSGIFPFGVAQLVNAIDQGV
TGKIAKETADLMMESPCKNPEILKERPELLPMDEGELQKIIAEVVLNAPESIVDYK
NGTKALGFLVQIMKRTAGKAPPRVNNELLILDLKG"
4127..6895
/gene="pmp_1"
CDS
4127..6895
/gene="pmp_1"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family"
/protein_id="BAA98215.1"
/db_xref="GI:8978378"
/translation="MRFSLCGFPLVFSFTLLSVFDTLSATTISLTTPEDSFHGDQNA
ERSYNVQAGDYSILTDGVSISNVDNSALNKAFCNVTSGSVTFAGNHGHLFNNISST
TKEGALCCQDPQATARFSGFTSLFISQSGDIEKEQGLYSKNAIMLLNNYVVFEEQN
QSKTKGGAISGANTIVIGNYDSVSFYQNAATFGGAIHSSGVLQIYVNAQAEIRFQNTA
KNKGLGALDGDIDIDQNAVYLFRENEALTATCKGGAACCLPTSGSSTPVPIVPS
DNKQIVERNHISIMGGGAIYARKLSISGGPTLFINNISYANSONLGAIAIDTGGTI
SLSAEKGTITFOGNKTSPLPLNGIHLQNAKFLKQARNYSIEFYDPIITSEADGTSI
LINQDPKNKEVYCTILFSGEKSILANDPRPKSTIPQNVNLISAGILVIKGAETVYSK
FTQSPGSHLVDLDTGLKASKEDIAITGLAIDISLSSSSSTAATAVINKANTANKOISVD
SIELISPTGAYEDLNRMSNTQFPFLSLEPAGGSVTVTAGDPLVPSPHYGFGGNMKL
AWTGTGNKVGFFWDKINYPKPEKGNLVPNILNAGVADVRSILMOVQETHASSLOTD
RGLWDIGTGNFPHVSASEDNIRYRHNHSGGYVLSVNNETTPKHVTSMAFSQLFSRDQY
```

gene 13434..15758  
 CDS /gene="CPJ0010"  
 13434..15758  
 /gene="CPJ0010"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BA98220.1"  
 /db\_xref="GI:8978383"  
 /translation="MLLLISGALFLGIPGLSAAISFGLIGLSALGVLMTSGLLC  
 LLVKREIPTVRPEIPEVSLAPSEPALQAQKTLAQLPKELDLTDIOEVFACLR

Query Match 19.7%; Score 76.4; DB 1; Length 300050;  
 Best Local Similarity 54.7%; Pred. No. 3e-10;  
 Matches 152; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
 QY 35 aatagctcttcgagctctcttcgcagcttcgccaattggcttcgcagggcgcagcat 94  
 Db 222381 AGTTGGCCCTTCGATTATGGGGCAGATCTTACCTGTTTGGGTGTAGAAGCAAAAAA 222322  
 QY 95 gctcaactcggccgattgctccacatgacatgagtgatggcatttggcccaaa 154  
 Db 222321 AGAGCAGCGGGAAGTATTATACACATAGATATCATGATGCCACTTTGTTCGAA 222262  
 QY 155 tttaactattggcgtccagttattgaaagtgtgagaaagcacacaaaggcatttggga 214  
 Db 222261 CTTTACTTTTGGTCAGGATCATGTGCTGCCATTAATAGATCTACGGATCTATTTTGA 222202  
 QY 215 ttgtcacatttggttacaacatctcttattgattgttgaaaccttggcaaaagctgtgc 274  
 Db 222201 AGTCCACGCTATGATTACAACTCTTTTGAATTTATAGAAAGTTTCGTTCTGCTG 222142  
 QY 275 ttctggtttacattcagtagagacatacaaaagata 312  
 Db 222141 GGATGAATTATAGTACACTTTTGAAGCTTCAGAGGATA 222104

RESULT 14  
 AF070943 849 bp DNA SYN 30-JUN-1998  
 LOCUS Expression vector pFL506 for Spinacia oleracea chloroplast  
 DEFINITION ribulose-phosphate 3-epimerase mature form.  
 ACCESSION AF070943  
 VERSION AF070943.1 GI:3264791  
 KEYWORDS Expression vector pFL506.  
 SOURCE Expression vector pFL506.  
 ORGANISM artificial sequence; vectors.  
 REFERENCE 1 (bases 1 to 849)  
 AUTHORS Larimer,F.W., Lu,T.-Y.S., and Stringer,C.D.  
 TITLE Construction, purification, and characterization of recombinant D-ribulose-5-phosphate 3-epimerase from Spinacia oleracea  
 JOURNAL Biochemistry 35, 9296 (1996)  
 REFERENCE 2 (bases 1 to 849)  
 AUTHORS Chen,Y.-R., Lu,T.-Y.S., Hartman,F.C. and Larimer,F.W.  
 TITLE Purification and properties of recombinant spinach ribulose-5-phosphate 3-epimerase  
 JOURNAL Plant Physiol. (1998) In press  
 REFERENCE 3 (bases 1 to 849)  
 AUTHORS Larimer,F.W. and Lu,T.-Y.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUN-1998) Life Sciences Division, Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, TN 37831, USA  
 FEATURES  
 source 1..849  
 /organism="Expression vector pFL506"  
 /db\_xref="taxon:78122"  
 /lab\_host="Escherichia coli"  
 /note="synthetic expression cassette for Spinacia oleracea chloroplast ribulose-phosphate 3-epimerase mature form"  
 1..849  
 /gene="RPE"  
 1..97

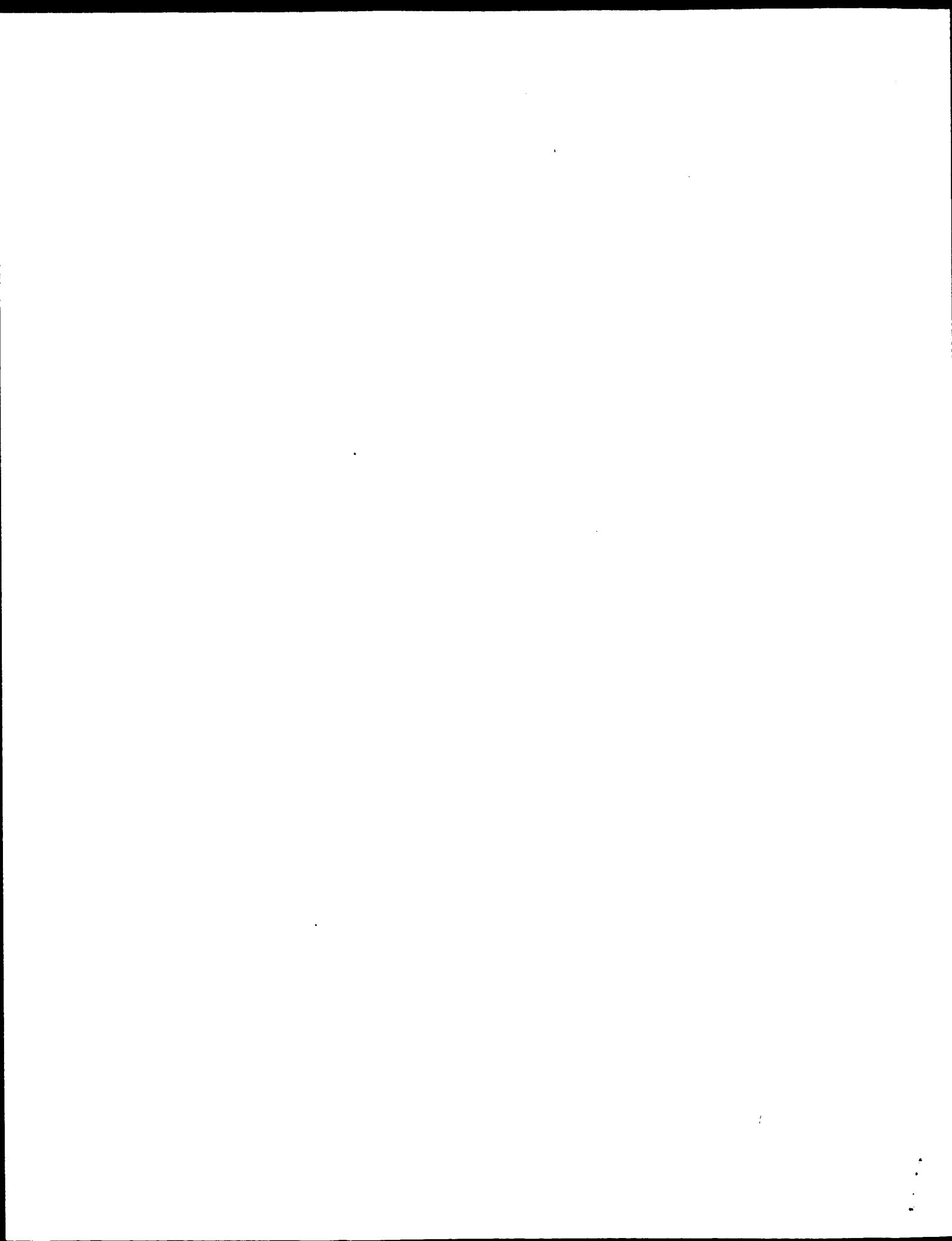
CDS  
 /gene="RPE"  
 /note="synthetic tac (trp-lac) promoter"  
 98..811  
 /gene="RPE"  
 /EC\_number="5.1.3.1"  
 /note="pentose-5-phosphate 3-epimerase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribulose-phosphate 3-epimerase mature form"  
 /protein\_id="AAC24710.1"  
 /db\_xref="GI:3264792"  
 /translation="MATSRVDKFKSDIIVSPSILSANFAKLGEOVKAVELAGCDWIH  
 VDMDGFEVNNITIGPLVVDALRPVTDLPDLVHLMIVPEQRVPDFIKAGADIYSVHC  
 ELASTIHLHRTVNOIKSLGAKAGVVLNPGTLPSTIEYVLDVLDVLMVNPFGGOS  
 FTESOVKKISDLRKMVEKGVNPNWIEVGGVTPANAKYVIEAGANALVAGSAVFGARD  
 YAEAIKGINASKRPEPVA"  
 BASE COUNT 233 a 168 c 210 g 238 t  
 ORIGIN  
 Query Match 19.4%; Score 75.4; DB 91; Length 849;  
 Best Local Similarity 53.0%; Pred. No. 4e-10;  
 Matches 185; Conservative 0; Mismatches 161; Indels 3; Gaps 1;  
 QY 40 ctcttcgagctctcttcgcagcttcgccaattggcttcgcagggcgcagcgtctcc 99  
 Db 147 CTCCATCTATTCTCTCTGTAACCTTTGCAAGTTGGGTGAGCAGGTAAAGCAGTTGAAT 206  
 QY 100 acttcggccgctgattggctccacatggacatcatggtggtgatttggcccaatttaa 159  
 Db 207 TGGCTGGTGTGATTGATTGATGTTGATGTAATGGATGGTGTTCCTTTCCTCAAAATATTA 266  
 QY 160 ctattggcgtccagttattgaaagtgtgagaaagcacacaaaggcatttggattgtc 219  
 Db 267 CTATTGGACCTCTGGTGTGATGTCCTTCGACCTGTGACATCTCCACATAGATGTC 326  
 QY 220 acctatggttacaacatctcttattgattgtgaaccttggcaaaccttggcaaaccttgc 279  
 Db 327 ATCTGATGATTGTGAGAACGACGAGCAGCGAGTCCAGACTTCATCAAGCTGGTGTGATA 386  
 QY 280 gttttacatttcagctagag--acatcaaaagataactggaaagaacttatccaaagaa 336  
 Db 387 TAGTGAGTGTCCACTGTGACTAGCATTCCACCATTCATTTCATCGACCGCTAAACCAAAA 446  
 QY 337 tcaagtcaacatggcagctgattctctgtgtgtagcattaaagccttggagcccc 385  
 Db 447 TCAGAGAGTTGGGAGCTAAAGCTGGAGTGTCTTAACCTGGAACCCC 495

RESULT 15  
 AF070942 993 bp DNA SYN 30-JUN-1998  
 LOCUS Expression vector pFL505 for Spinacia oleracea chloroplast  
 DEFINITION ribulose-phosphate 3-epimerase transit form.  
 ACCESSION AF070942  
 VERSION AF070942.1 GI:3264789  
 KEYWORDS Expression vector pFL505.  
 SOURCE Expression vector pFL505.  
 ORGANISM artificial sequence; vectors.  
 REFERENCE 1 (bases 1 to 993)  
 AUTHORS Larimer,F.W., Lu,T.-Y.S., and Stringer,C.D.  
 TITLE Construction, purification, and characterization of recombinant D-ribulose-5-phosphate 3-epimerase from Spinacia oleracea  
 JOURNAL Biochemistry 35, 9296 (1996)  
 REFERENCE 2 (bases 1 to 993)  
 AUTHORS Chen,Y.-R., Lu,T.-Y.S., Hartman,F.C. and Larimer,F.W..  
 TITLE Purification and properties of recombinant spinach ribulose-5-phosphate 3-epimerase  
 JOURNAL Plant Physiol. (1998) In press  
 REFERENCE 3 (bases 1 to 993)  
 AUTHORS Larimer,F.W. and Lu,T.-Y.S.  
 TITLE Direct Submission

JOURNAL	Submitted (06-JUN-1998)	Life Sciences Division, Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, TN 37831, USA
FEATURES	Location/Qualifiers	
source	1. .993	
	/organism="Expression vector pFL505"	
	/db_xref="taxon:78121"	
	/lab_host="Escherichia coli"	
	/note="synthetic expression cassette for Spinacia oleracea chloroplast ribulose-phosphate 3-epimerase transit form"	
gene	1. .993	
	/gene="RPE"	
promoter	1. 97	
	/gene="RPE"	
	/note="synthetic tac (trp-lac) promoter"	
CDS	98. .955	
	/gene="RPE"	
	/EC_number="5.1.3.1"	
	/note="pentose-5-phosphate 3-epimerase"	
	/codon_start=1	
	/transl_table=11	
	/product="ribulose-phosphate 3-epimerase transit form"	
	/protein_id="AAC24709.1"	
	/db_xref="GI:3264790"	
	/translation="MGAASLCSTQSQINGCGGKINIKLPSTSSPNSLFTTRKKV QTIVKATSVDFKSKSDIIVSPSILSANFAKLGQKAVELAGCDWIHVDVMDGRVP NTITGLVDALRPVTDLPDLVHLMTIVEPEQRPVDFIKAGADIVSVHCELASTIHLHR TVNQIKSLGAGVIVPDLPTLSTIEYLDVLDVLIMSVNPFGGQSFTISQGVKKIS DURKMCVEKGVNPIEVDGGVTPANAYKVEAGANALVAGSAVFAGAKDYAEAIKGIKA SKRPEPVAY"	
BASE COUNT	268 a	202 c 237 g 286 t
ORIGIN		

Query Match	19.4%	Score 75.4;	DB 91;	Length 993;
Best Local Similarity	53.0%;	Pred. No. 4e-10;		
Matches 185;	Conservative 0;	Mismatches 161;	Indels 3;	Gaps 1;
QY	40	ctcttcgatgctctcttcgcagacttcgcgaatttggcttcaggagctcaagcagatgctcc	99	
DB	291	CTCATCTATTCTCTCTGCTTAACCTTTGCAAGATGGGTGACGAGGTAAACAGCATTTGAAT	350	
QY	100	acttcggcgccgattggctccacatgagacatcatgatgatggcattttgtccccaatttaa	159	
DB	351	TGGCTGGTTGTGATGGATTCATGTTGATGATGATGATGCTGTTTGTTCCAAAATATTA	410	
QY	160	ctattggcgctccagttattgaaagtttgagaaagcacacaaagggcatatttggattgtc	219	
DB	411	CTATTGGACCTCTGGTGGTTTGATGTCCTTGGGACCTGTGACAGATCTCCCACTAGATGTC	470	
QY	220	accttatggttacaatctcttggattattgttaaccttggcacaagctgggtgctctgt	279	
DB	471	ATCTGATGATTGTAGAACCCAGACGCGGATCCCGAGCTTCATCAAGCTGGTGCCTGATA	530	
QY	280	gttttacatttcacgtatag- --acatcaaaagataactggaaagaacttatccaaagaa	336	
DB	531	TAGTGAGTGTCCACTGTGAGCTAGCATCCACCATTCATTTGCATCGGACCGTAAACCAA	590	
QY	337	tcaagtcacatggcagtgattcttcgttgtagcattaaagctcgggagcccc	385	
DB	591	TCAAGAGTTTGGGAGCTTAAAGCTGGAGTTGTCTTTAAACCTGTGAACCC	639	

Search completed: November 4, 2000, 13:36:23  
Job time: 17232 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:01 ; Search time 320.8 Seconds  
(without alignments)  
454.355 Million cell updates/sec

Title: US-09-300-482-298  
Perfect score: 388  
Sequence: 1 ggaagaagaagaaagatg.....ttaaaagcctggagccccccgt 388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues  
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : N\_Geneseq\_36 : \*
- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:\*
  - 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:\*
  - 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:\*
  - 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:\*
  - 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:\*
  - 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:\*
  - 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:\*
  - 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:\*
  - 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:\*
  - 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:\*
  - 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:\*
  - 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:\*
  - 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:\*
  - 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:\*
  - 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:\*
  - 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:\*
  - 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:\*
  - 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:\*
  - 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:\*
  - 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:\*
  - 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
C 1	83.6	21.5	18 V75460	Staphylococcus aur
C 2	77.8	20.1	17082 20 X13166	Enterococcus faeca
C 3	77.5	20.0	12566 19 V52282	Streptococcus pneu
C 4	74.2	19.1	32768 20 X20515	Polynucleotide seq
5	69	17.8	1050 21 X45413	D-ribulose-5-phosp
6	65.8	17.0	1260 21 X45412	D-ribulose-5-phosp
7	65.4	16.9	1234 19 T98644	DNA encoding a S.
8	65.4	16.9	1542 19 T96346	S. pneumoniae deri
C 9	58	14.9	1019 21 A26928	Essential staphylo
10	53.2	13.7	1040 19 V65230	DNA encoding a S.
11	43.6	11.2	459 21 A31245	Plant microsatelli
12	41.8	10.8	11443 19 V52182	Streptococcus pneu

C 13	38.2	9.8	6866 13	Q27406	Human parainfluenz
14	37.6	9.7	439 21	A31286	Plant microsatelli
15	36.4	9.4	573 20	X61720	B. burgdorferi ant
16	36.4	9.4	606 20	X61719	B. burgdorferi ant
C 17	34	8.8	3781 19	V26296	Phosphate starvati
18	33.6	8.7	434 21	A31266	Plant microsatelli
19	33.4	8.6	1222 21	V25673	Human transmembran
20	33.2	8.6	1486 19	V71042	CDNA encoding (iso
21	32.2	8.3	4741 19	V23686	L.lactis MG1363 Ad
C 22	32	8.2	20633 20	X13213	Enterococcus faeca
C 23	31.8	8.2	1380 21	A26697	Enterococcus faeca
C 24	31.6	8.1	10555 20	X13170	Candida albicans p
C 25	31.6	8.1	10749 20	X20257	Borrelia burgdorfe
C 26	31.6	8.1	40352 19	V02032	MAGE-B cluster DNA
27	31.4	8.1	7646 18	V74432	Staphylococcus aur
28	31.2	8.0	1327 19	X14338	H. pylori GHPO 236
29	31.2	8.0	1440 19	V65769	Campylobacter jeju
C 30	31	8.0	1430 16	Q92525	Nicotiana plumbagi
31	30.8	7.9	2290 19	V59677	Human secreted pro
C 32	30.6	7.9	506 21	V57037	PBR-associated pro
C 33	30.6	7.9	3150 19	V05542	Actinobacillus ple
34	30.4	7.8	969 20	Z07815	S. aureus nrdf pol
35	30.4	7.8	5897 18	V74831	Staphylococcus aur
36	30.4	7.8	580073 18	T58840	Mycoplasma genital
C 37	30.4	7.8	1230025 20	X91990	Nucleotide sequenc
38	29.8	7.7	2622 18	T90461	Nucleic acid seque
39	29.8	7.7	2622 20	Z19861	Thermoanaerobacter
40	29.8	7.7	2622 20	Z19861	Thermoanaerobacter
41	29.8	7.7	4617 17	T72553	Rat tripeptidylpep
42	29.8	7.7	4630 20	V84139	Human prothrombina
43	29.8	7.7	5109 17	T72654	Rat brain homoloqu
44	29.8	7.7	5300 18	T90447	Thermoanaerobacter
45	29.8	7.7	5300 20	Z19859	Thermoanaerobacter

ALIGNMENTS

RESULT 1	
V75460/c	
ID V75460 standard; DNA; 916 BP.	
XX	
AC V75460;	
XX	
DT 16-MAR-1999 (first entry)	
XX	
DE Staphylococcus aureus contig SEQ ID #1149.	
XX	
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;	
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;	
KW skin infection; surgical wound infection; scalded skin syndrome;	
KW toxic shock syndrome; ds.	
XX	
OS Staphylococcus aureus.	
XX	
Key Location/Qualifiers	
misc_feature 1..60	
FT /tag= a	
FT /note= *these bases represent a line of missing text in	
FT the sequence listing in the specification. They	
FT are included to maintain the nucleotide numbering	
FT given in the specification for this DNA sequence*	
EP786519-A2.	
XX	
PD 30-JUL-1997.	
XX	
PF 07-JAN-1997; 97EP-0100117.	
XX	
PR 05-JAN-1996; 96US-0009861.	
XX	
PA (HUMA-) HUMAN GENOME SCI INC.	
XX	

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
XX Rosen CA;  
XX WPI; 1997-374922/35.  
XX  
XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
XX  
XX  
XX Claim 1; Page 1854-1855; 3271pp; English.  
XX  
XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
XX  
XX Sequence 916 BP; 273 A; 165 C; 120 G; 291 T; 67 other;  
SQ  
  
Query Match 21.5%; Score 83.6; DB 18; Length 916;  
Best Local Similarity 53.6%; Pred. No. 4.8e-17;  
Matches 192; Conservative 2; Mismatches 161; Indels 3; Gaps 1;  
  
Qy 31 cgaataagctcctcgatgctctccgacttcgccaatttggttcgcgaggtcagc 90  
Db 745 CAAACATATATCCATCAATATATATCTGTTGATTTTGGATTACACATGATTAAC 686  
  
Qy 91 gcatgtccacttcggcgcgattggctccacatggacatcatggtggtggttccttc 150  
Db 685 GACITGGAAGAGCAGCGCTGACGGAGTTCATTTTGTATGTTGATGGTCAATTTGTGC 626  
  
Qy 151 ccaatttaactattggcgcctccagttattgaagtttgaaagacacacaaagggcatatt 210  
Db 625 CTAATATATCTATTTGGTTTACCAATATATAGATGATGAGTAAGAAAGGCACACATTA 566  
  
Qy 211 tggattgtcacttatggttcaaatcctcttgattatgttgaaaccttggcgaagctg 270  
Db 565 TAGACGTACATTTGATGATTAATCCAGAAAGATATTTGTCATCATTTGCAGAACATG 506  
  
Qy 271 gtgcttctgttttacatttcacgttagacatcaacaaagataactggaagaaacttatcc 330  
Db 505 GTGCGGATATGATTCATATTCATCTCGA---ATCAAGCCCTCATATTCATGCTGCTATTC 449  
  
Qy 331 aaagatacaagtcacatggcatgattcctggtgtagcattaaagcctgggaccccggt 388  
Db 448 AAATGATTAACATTTAGATAAANAAGCTGGTGTAGTAATTAATCCCTGGTACMCCMAT 391  
  
RESULT 2  
X13166/C  
ID X13166 standard; DNA; 17082 BP.  
XX  
XX X13166;  
AC  
XX  
XX 19-MAR-1999 (first entry)  
DT  
XX  
XX Enterococcus faecalis genome contig SEQ ID NO:229.  
DE  
XX  
XX Enterococcus faecalis; contig; detection; Enterococcal infection;  
KW

KW vaccine; attenuation; computer readable medium; ds.  
XX Enterococcus faecalis.  
XX WO9850555-A2.  
XX  
XX 12-NOV-1998.  
XX  
XX 04-MAY-1998; 98WO-US08985.  
XX  
XX 14-NOV-1997; 97US-0066009.  
XX  
XX 06-MAY-1997; 97US-0044031.  
XX  
XX 16-MAY-1997; 97US-0046655.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Barash SC, Dillon PJ, Kunsch CA;  
XX WPI; 1999-045171/04.  
XX  
XX New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
XX  
XX Claim 1; Page 1149-1157; 2084pp; English.  
XX  
XX A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
XX  
XX Sequence 17082 BP; 4548 A; 3830 C; 3167 G; 5533 T; 4 other;  
SQ  
  
Query Match 20.1%; Score 77.8; DB 20; Length 17082;  
Best Local Similarity 52.3%; Pred. No. 1.5e-14;  
Matches 197; Conservative 0; Mismatches 177; Indels 3; Gaps 1;  
  
Qy 9 aaagaaaagatgggaatgacacccgaaatagctccacttcggttcggttccttcgacttcgc 68  
Db 7372 AATATAAGGGGAGCGAGCCAAATGAACATAGCACCATCCATTTTAAAGTCCGATTTTGGC 7313  
  
Qy 69 aatttggtctccgaggtcagcgcatgctccacttcggttcggttcggttcggttcggttcg 128  
Db 7312 AATTAGAAAGAGATATCCGTTTAGTAGAAGAAATAGGAGCAGATATATTCATGTGTAT 7253  
  
Qy 129 atcatggatgggcatcttgcctcccaatttaactattggcggtccagttattgaagtttg 188  
Db 7252 GTCATGACGGTCAATTTGTGCCCAACATATACCTTAGGACCGAATATCCTTTACGGATT 7193  
  
Qy 189 aagaagcacacaaggcgcataatttggttgatgtccacttcgtttacaaactcctcttgatt 248  
Db 7192 CGTCCAGTGACAAAACATACCATTTGGATGTTTCATTTTGTATGATTTGTACAACTGAAATAT 7133  
  
Qy 249 gttgaaccttggcaaaagctggtgtcttctggtttttacatttcacgttagagacatacaaa 308  
Db 7132 ATCAATGCTTTTGGGAAGCGCGTGGGATATATTTACCTGCCATCAAGAAGCAACACC 7073  
  
Qy 309 gataactggaaaacttatccaaagaactcaagtcacatggcatgattcctcgtgttagca 368  
Db 7072 CATA---TTCATCGTCCTTACAAATGATTAATAATGCGGGTGTGAAGCTGCGGTANCC 7016







DR WPI; 2000-097543/08.  
 XX P-PSDB; Y54150.  
 XX Nucleic acid encoding D-ribulose-5-phosphate-3-epimerase is used to  
 PT increase resistance to root nematodes -  
 XX Claim 1; Page 44-45; 62pp; French.  
 XX The present sequence encodes an Arabidopsis D-ribulose-5-phosphate-  
 CC 3-epimerase (RPE) nematode response protein. The enzyme is activated  
 CC during the initial stages of nematode infestation of roots and is  
 CC required for formation of the giant feeder cells on which the nematodes  
 CC depend. Interfering with expression of the protein will thus inhibit  
 CC nematode infestation. The RPE polynucleotide is used to produce  
 CC transgenic plants (specifically potato, tomato, beet, rape and rice)  
 CC that are resistant (or at least less susceptible) to nematode  
 CC infestation. This is achieved by either expressing antisense sequences  
 CC in the plant; or by using the promoter of the RPE gene to express an  
 CC anti-nematode protein. Inhibition of the RPE protein or gene enables  
 CC plant to mount an immediate response to the invasion of the nematode,  
 CC consequently limiting the level of cellular damage and reducing the  
 CC complications of secondary infections.  
 XX Sequence 1260 BP; 365 A; 236 C; 288 G; 371 T; 0 other;

Query Match 17.0%; Score 65.8; DB 21; Length 1260;  
 Best Local Similarity 51.3%; Pred. No. 2.7e-11;  
 Matches 179; Conservative 0; Mismatches 167; Indels 3; Gaps 1;  
 QY 40 ctcttcgatgctctcttcgcagcttcgccaatttgcttcaggctcagcgatgctcc 99  
 Db 287 ctccctattctcgtgtaatttcgcaattagcgagcaggtaaaagcagtgga 346  
 QY 100 acttcggcgcgattggtccacatgcagatcatggtggtgctttgtccccaattaa 159  
 Db 347 tggcagggttgattgattcattgtgatgcaggcgtggtttgttcccaacatta 406  
 QY 160 ctattggcgtccagttattgaagttgagaagacacacaaaggcatttggattgctc 219  
 Db 407 ctatcgagacctctcggttgatgctcttgcgcttgacagatctcttctggatgttc 466  
 QY 220 acctatggttacaaatctcttgattgttgaaccttggcaaaagctggtctctg 279  
 Db 467 atctgatgagtggaaccgcagacagagatccggtatttcacaaagcagtgagata 526  
 QY 280 gttttacattca---cgtagacatcaaaagataactggaagaaacttaccaaagaa 336  
 Db 527 ttgttagtgatattgtaacagcaatccaccatccatttgcacgtaccgtcaatcaaa 586  
 QY 337 tcaagtcacatgcagatgattcctggtgtagcattaaagcgtggacc 385  
 Db 587 taaaagcttagggctaaagctggaggtgttctcaaaccttggaaaccc 635

RESULT 7  
 T98644  
 ID T98644 standard; DNA; 1234 BP.  
 XX T98644;  
 AC T98644;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE DNA encoding a S. pneumoniae protein of unknown function.  
 XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
 KW immunological response; inoculation; antibody production; inhibitor;  
 KW T cell immune response; antimicrobial compound; bacterial adhesion;  
 KW extracellular matrix protein; protein-mediated cell invasion; wound;  
 KW pathogenesis; ss.  
 XX Streptococcus pneumoniae.  
 OS  
 XX

FH Key Location/Qualifiers  
 FT CDS 55..933  
 XX /\*tag= a  
 PN WO9743303-A1.  
 XX 20-NOV-1997.  
 XX  
 PF 14-MAY-1997; 97WO-US07950.  
 XX  
 PR 14-MAY-1996; 96US-0017670.  
 XX  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Stodola RK;  
 XX  
 DR WPI; 1998-008793/01.  
 DR P-PSDB; W38594.  
 XX  
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for  
 PT diagnosing anti-microbial agents for treatment of bacterial  
 PT infections  
 XX  
 PS Claim 4; Pages 167-168; 483pp; English.  
 XX This sequence encodes a Streptococcus pneumoniae protein of unknown  
 CC function, and represents a DNA sequence of the invention.  
 CC The DNA sequences are isolated from Streptococcus pneumoniae strain  
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the  
 CC invention can be used to identify compounds which interact with and  
 CC inhibit or activate the activity of the proteins. Antagonists can be  
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic  
 CC immunisation. They can also be used to induce an immunological response  
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery  
 CC of the encoding nucleic acids in a vector adequate to produce antibody  
 CC and/or T cell immune responses to protect the animal from disease. The  
 CC proteins can also be used to identify antimicrobial compounds which are  
 CC capable of inhibiting their bioactivity. In particular the proteins of  
 CC the invention can be used to prevent adhesion of bacteria to mammalian  
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
 CC block protein-mediated mammalian cell invasion, and to block the normal  
 CC progression of pathogenesis in infections initiated other than by the  
 CC implantation of in-dwelling devices or other surgical techniques.  
 XX  
 SQ Sequence 1234 BP; 355 A; 260 C; 278 G; 341 T; 0 other;  
 Query Match 16.9%; Score 65.4; DB 19; Length 1234;  
 Best Local Similarity 55.5%; Pred. No. 3.6e-11;  
 Matches 126; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
 QY 33 aaaatagctcttcgatgctctcttcgcagcttcgccaatttgcttcgaggtcagcgc 92  
 Db 956 aagatgctcgtcaattctctggcagcagattatgccaaacttggaactcaaacgt 1015  
 QY 93 atgtccacttcggcgccgattggtccacatggacatcatggatggcatttggctccc 152  
 Db 1016 ctagaagcaactggggcagaatattgccatattgcacatcatgacatgacatgacacg 1075  
 QY 153 aatttaactattgctgctccagttattgaaagtttgaaaagcagacaaaggcatttg 212  
 Db 1076 caaatcagtttgggtgcaggtggtgagagccttgcctcatagtaagatggttttc 1135  
 QY 213 gatgtcaccttatggtttacaaatcccttgcatttatgtatgaacccctt 259  
 Db 1136 gattgccaacttgatgggtgtcaaaccttgaaacatcatcatctcggaaaaatt 1182  
 RESULT 8  
 T96346/c  
 ID T96346 standard; DNA; 1542 BP.

XX Z96346;  
AC 10-APR-2000 (first entry)  
DT S. pneumoniae derived DNA from ORF #174.  
DE Treatment; prevention; disease; diagnosis; gene therapy; screening;  
XX bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.  
KW Streptococcus pneumoniae.  
XX WO9806734-A1.  
XX 19-FEB-1998.  
PD 15-AUG-1997; 97WO-US14436.  
XX 16-AUG-1996; 96US-0024022.  
XX (SMK ) SMITHKLINE BEECHAM CORP.  
PA Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
PI Stodola RK;  
XX WPI; 1998-159452/14.  
DR P-PSDB; Y86006.  
XX Streptococcus pneumoniae proteins and related DNA - useful for  
PT screening compounds for antibacterial activity  
XX Claim 4; Page 207-208; 640pp; English.  
XX This invention describes novel isolated Streptococcus pneumoniae  
CC polynucleotides (see 236173-236494) and their encoded proteins (see  
CC Y85792-Y86182). The DNA, vectors and host cells described in the method  
CC of the invention are useful for the recombinant expression of the  
CC polypeptides. The polypeptides are useful for treatment or prevention of  
CC disease, or diagnosis of disease related to expression or activity of  
CC such a polypeptide. They can also be used to screen for compounds which  
CC interact with and inhibit or activate such a polypeptide. The  
CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
CC for inducing an immunological response in a mammal. The antagonists are  
CC useful to inhibit such bacterial polypeptides. The polypeptides are  
CC particularly useful to identify antimicrobial compounds and antibiotics.  
CC They are also useful to determine their role in pathogenesis of  
CC infection, dysfunction and disease.  
XX Sequence 1542 BP; 421 A; 367 C; 328 G; 426 T; 0 other;

Query Match 16.9%; Score 65.4; DB 19; Length 1542;  
Best Local Similarity 55.5%; Pred. No. 4e-11;  
Matches 126; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 33 aaaaatagcttccttcgactctctccgacttcgcaatttggttcogaggtcagcgc 92  
DB 279 AAGATTGCTCCGTCATTCGGCAGCAGATATGCGCACTTGAACGGAATCAACGT 220  
QY 93 atgtccacttcggcgccgattggtccacatggacatggatggatggatggatggat 152  
DB 219 CTAGAAGCAACTGGGGCAGAAATATGCCATATCGATATCATGGACAGTCATTTGTACCG 160  
QY 153 aatttaactattggcctccagttattgaagtttgagaagacacacaaagccatttg 212  
DB 159 CAATCAGTTTGGTGGAGGTGGTGGAGAGCCCTCGTCCTCATAGTAAGATGGTTTC 100  
QY 213 gattgtcaccttatggttacaaatcctcttgattatgttgaaacctt 259  
DB 99 GATTGCCACTTGGTGTCAAAACCTTGAGCATCATCTCTGGAAGATT 53

RESULT 9

A26928/C  
ID A26928 standard; DNA; 1019 BP.  
XX AC A26928;  
XX 29-JUN-2000 (first entry)  
XX Essential Staphylococcus aureus gene #79.  
XX Essential gene; Staphylococcus aureus infection; screening; prevention;  
KW antibacterial agent; bacterial infection; ss.  
XX Staphylococcus aureus.  
XX US6037123-A.  
XX 14-MAR-2000.  
XX 13-SEP-1996; 96US-0714918.  
XX 15-SEP-1995; 95US-0003798.  
XX 22-DEC-1995; 95US-0009102.  
XX (MICR-) MICROCID PHARM INC.  
XX Benton B, Lee VJ, Martin PK, Schmid MB, Sun D, Malouin F;  
XX WPI; 2000-282222/24.  
XX Novel methods for screening for antibacterial agents, useful for  
PT treatment or prevention of Staphylococcus infection, by testing  
PT compounds against, essential bacterial genes -  
XX Disclosure; Column 183-184; 256pp; English.  
XX The present sequence represents an essential Staphylococcus aureus gene.  
CC The invention relates to a method for screening for an antibacterial  
CC agent. The method comprises determining if a test compound is active  
CC against the essential bacterial genes (A26850-A26956). The invention also  
CC includes a method for evaluating an agent for activity on the essential  
CC genes. Also included in the invention is the production of an  
CC antibacterial agent. The antibacterial agents of the invention are used  
CC to treat or prevent bacterial infections, particularly where caused by  
CC Staphylococcus aureus. The antibacterial agents are unlikely to be  
CC affected by known resistance mechanisms.  
XX Sequence 1019 BP; 314 A; 211 C; 151 G; 340 T; 3 other;

Query Match 14.9%; Score 58; DB 21; Length 1019;  
Best Local Similarity 56.9%; Pred. No. 7.6e-09;  
Matches 161; Conservative 0; Mismatches 118; Indels 4; Gaps 3;  
QY 105 ggcgcgattgggtccacatggacatggatggatggatggatggatggatggatggat 164  
DB 1001 GGNATTCACGAGTTCATTTTGATGTTGATGGATGGTCAATTTGCCCTAATATATATT 942  
QY 165 ggcgcctccatttgaagtttgagaagacacacaaagccatttgatggatggatggat 224  
DB 941 GGTTTACCAATATAGATGAGTAAGAAAGGCGCACACATTAACNTATAGACGTACATTG 882  
QY 225 atggtt-acaaatcctcttgattatgttgaaaccttggaacacacacacacacacacac 282  
DB 881 ATGATTGAAAATCCAGAAAGATATATGATCATCATTTGCGAGACATGGTGGCGATATGAT 822  
QY 283 ttacatttcacgtagac 342  
DB 821 TTNCAATTCATGCGAATCAACGCCCTCATA--TTTCATCGTCTATTTCAATGATTAAC 764  
QY 343 cacatggcatgattcctctggttagcattaaagcctggagccccc 385  
DB 763 ATTAGATAAAAAAGCTGGTGTAGTAATTAATCTCGGTACACC 721

```

RESULT 10
V65230
ID V65230 standard; DNA; 1040 BP.
XX
AC V65230;
XX
DT
XX
DE
XX
DE DNA encoding a S. pneumoniae protein of unknown function.
XX
KW Streptococcus pneumoniae protein; recombinant; gene expression;
KW DNA chip; virulence; antibody; infection; detection; treatment; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO9826072-A1.
XX
PD 18-JUN-1998.
XX
PF 09-DEC-1997; 97WO-US22578.
XX
PR 13-DEC-1996; 96US-0036281.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
PI Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR;
PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;
PI Young Bellido ML;
XX
XX WPI; 1998-348529/30.
XX
DR P-PSDB; W80657.
XX
PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT for evaluating gene expression, and identification of virulence
PT genes
XX
XX Claim 1; Page 75; 333pp; English.
XX
CC This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function. The invention provides DNA sequences (V65201 to V65304) from
CC the Streptococcus pneumoniae genome and corresponding protein sequences
CC (W80605 to W80728). A recombinant host containing a vector comprising any
CC of the above nucleic acids can be used for the recombinant expression of
CC the protein sequences. The invention also provides a DNA chip having
CC arrayed on it at least 15 base pair fragment of any one or more of these
CC DNA sequences. The DNA chip can be used methods for evaluating gene
CC expression in S. pneumoniae and for identifying virulence genes in
CC S. pneumoniae. Antibodies that selectively bind to the above proteins or
CC peptide fragments can be used to treat S. pneumoniae infection. The
CC antibodies can also be used to detect S. pneumoniae cells.
XX
SQ Sequence 1040 BP; 292 A; 236 C; 227 G; 285 T; 0 other;

Query Match 13.7%; Score 53.2; DB 19; Length 1040;
Best Local Similarity 52.5%; Pred. No. 2.6e-07;
Matches 187; Conservative 0; Mismatches 163; Indels 6; Gaps 3;

QY 33 aaatagctccttcgatgctctcttcgcacttcgcacatttgcttcgaggtcagcgc 92
Db || || || || || || || || || || || || || || || || || || || || ||
652 aagattgctccgtcaattctggcagacattatgcacactttggaacgtgaaatacaact 711

QY 93 atgctccacttcggcgccgattggctccacatggacatcattggatggcatttgcgcc 152
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
712 ctagaagcaactggggcagaatcgcacatcgc--attctggacagtcatttgcacc 769

QY 153 aattcaactattggcgtcccatatttgaagatttgaaagacacacaaagcattttg 212
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
770 caaatcagtttggcgcagggtgggtccagagctt-cgtctcctcatgtaagatggtttc 828

QY 213 gattgtcacttatggttacaaactcctcttattattgttgaccccttggaagagctggt 272

```

```

Db 829 gattgccacttgatgtgtcaaacctcagcatcatctggaagattttgcgctgcaggt 888
QY 273 gcttctggttttacatttcacgttagacacatcaaaaagataactgaaaagaacttatccaa 332
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
889 gcagacatcatcagtcattccatgtagaagaacacacctcatattcattggtgcgcc---tccaa 945

QY 333 agaataagtcacatggtcattgattcctctggttagcattaaagcctgggaccccgct 388
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
946 aaattcgttcactcggagtttaagccttcagtcgttatcaatcctcctggcacaccagt 1001

RESULT 11
A31245
ID A31245 standard; DNA; 459 BP.
XX
AC A31245;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #206.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
XX Eucalyptus grandis.
XX
OS WO9967421-A1.
XX
PN 29-DEC-1999.
XX
PD 25-JUN-1999; 99WO-NZ00092.
XX
PF 25-JUN-1998; 98US-0105307.
XX
PR (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukkala IJ, Bloksberg LN, Glenn M;
XX
XX WPI; 2000-116958/10.
XX
DR New plant microsatellite markers and associated flanking species.for
DR the detection of polymorphic genetic markers -
XX
XX Claim 1; Page 136; 392pp; English.
XX
CC Sequences A31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences A32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety and
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 459 BP; 121 A; 113 C; 120 G; 103 T; 2 other;

```

```

Query Match 11.2%; Score 43.6; DB 21; Length 459;
Best Local Similarity 58.5%; Pred. No. 0.0002;
Matches 76; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 37 tagctccttcgatgctctcttcgcacttcgcacatttgcttcgaggtcagcgcagtc 96
Db || || || || || || || || || || || || || || || || || || || || ||
329 tatctccatcaattctatctgctaatttttcgaagctgggagatcaggtgaaagctgtgg 388

QY 97 tccacttcggcgcgatttggtcctcacatgagatcattggtggcatttgcctcccaatt 156

```

Db 389 agttggcaggtgattggtacacgttgatgtaagtgatgacggtttgtttcccaata 448  
 QY 157 taactattgg 166  
 Db 449 ttacaatcgg 458

## RESULT 12

V52182

ID V52182 standard; DNA; 11443 BP.

XX AC V52182;

XX 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:49.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

XX computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19588.

XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

XX Kunsch CA, Rosen CA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

XX polynucleotide sequences - useful in diagnostic kits and assays, and

XX pharmaceutical compositions and vaccines for Streptococcus

XX pneumoniae

XX Claim 1; Page 445-451; 1409pp; English.

XX The present invention describes a computer readable medium which has

XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded

XX on it, or a representative fragment or a sequence at least 95% identical

XX to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1

XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus

XX pneumoniae. The present invention also describes an isolated nucleic acid

XX molecule encoding a homologue of any of the fragments of the S. pneumoniae

XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced

XX by a process comprising: (a) screening a genomic DNA library using as a

XX probe a target sequence defined by any of the sequences in SEQ ID NO:1

XX to 391, identifying members of the library which contain sequences

XX that hybridize to the target sequence and isolating the nucleic acid

XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced

XX from an organism, amplifying nucleic acid molecules whose nucleotide

XX sequence is homologous to amplification primers derived from the

XX fragment of the S. pneumoniae genome to prime the amplification and

XX isolating the amplified sequences. The computer readable medium can be

XX used in a computer-based system for identifying fragments of the

XX S. pneumoniae genome of commercial importance, or expression modulating

XX fragments of the S. pneumoniae genome. Products from the present

XX invention can be used in diagnosis kits and assays, and pharmaceutical

XX compositions and vaccines for S. pneumoniae.

XX Sequence 11443 BP; 3350 A; 2057 C; 2533 G; 3502 T; 1 other;

XX Query Match 10.8%; Score 41.8; DB 19; Length 11443;

XX Best Local Similarity 47.2%; Pred. No. 0.0037;

Matches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
 QY 120 cacatggacatcatgatggcgcatattttgtccccaatttaactatttgccgctccagttatt 179  
 Db 7739 catatcgattattgatggcgcatattttgtccccaatttaactatttgccgctccagttatt 7798  
 QY 180 gaaagtttgagaaagcacacaaagccatattttgattgtcaccttatgtttacaaaatcct 239  
 Db 7799 caagaagtccaataattagtagcacacacatttatcagttcatctgatgtgcacagacca 7858  
 QY 240 ctgtattatgttgaaoccttggcaaaagctggtgttcttggttttacattttacacgtagag 299  
 Db 7859 accttttgggtagatcaagttctctgatatttacaatgtgagatattttgtattcatctgtaa 7918  
 QY 300 acatcaaaagataactggaaagaacttatccaaagaatcaagtcacatggcattgattcct 359  
 Db 7919 gttctgaatggtctgtctgttttggattgataaaattcatgatgcagggtctcaaggct 7978  
 QY 360 ggtgtagcattataagcctgggaccccgct 388  
 Db 7979 ggtgtgtcttaactcctgaaacacctgt 8007

## RESULT 13

Q27406/C

ID Q27406 standard; DNA; 6866 BP.

XX AC Q27406;

XX 05-FEB-1993 (first entry)

XX Human parainfluenza virus 2 large protein gene.

XX PIV-2; vaccine; ss.

XX Human parainfluenza virus-2.

XX JP04197179-A.

XX 16-JUL-1992.

XX 29-NOV-1990; 90JP-0325169.

XX 29-NOV-1990; 90JP-0325169.

XX (FUJG ) FUJIKURA KASEI KK.

XX WPI; 1992-288437/35.

XX Human para-influenza type 2 virus large protein gene - useful in

XX diagnosis and as a vaccine against para-influenza virus

XX Claim 1; Page 1-4; 17pp; Japanese.

XX Vero cells were cultured in the presence of actinomycin D and

XX infected with PIV-2. Poly(A)+ RNA was isolated from the infected

XX cells and a cDNA library was prepared. The library was screened with

XX a nucleocapsid RNA probe. A positive clone containing the PIV-2

XX gene was sequenced. The gene can be used for early diagnosis of

XX PIV-2 infection and as a raw material for vaccines.

XX Sequence 6866 BP; 2257 A; 1274 C; 1194 G; 2141 T; 0 other;

XX Query Match 9.8%; Score 38.2; DB 13; Length 6866;

XX Best Local Similarity 55.7%; Pred. No. 0.04; Mismatches 0; Gaps 0;

XX Matches 73; Conservative 0

QY 209 ttgtgattgtcaccttatgtttacaaatcctcttattgtttgaaccccttgcaaaagc 268

Db 1971 TTTTGTCTTTCCCTCTTATTTTAGAATCAGTCTTGATTGTGGAACCTTGGATGAGA 1912

QY 269 tgggtctctgtttttacatttcacgtagacatcaaaagataactggaaagaacttat 328

```

Db 1911 TGYTATCTCGGATATCTCTTCGCTTTCTGATATTATGCAATTTGACTCATAGTAA 1852
QY 329 ccaagaatca 339
Db 1851 GCAATGATTTA 1841

RESULT 14
A31286
ID A31286 standard; DNA; 439 BP.
AC A31286;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #247.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
PN WO9967421-A1.
XX
PD 29-DEC-1999.
XX
PF 25-JUN-1999; 99WO-NZ00092.
XX
PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR WPI; 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
XX
PS Claim 1; Page 148; 392pp; English.
XX
CC Sequences A31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences A32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 439 BP; 122 A; 95 C; 119 G; 102 T; 1 other;

Query Match 9.7%; Score 37.6; DB 21; Length 439;
Best Local Similarity 60.9%; Pred. No. 0.016;
Matches 78; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 40 ctcttcgatgctctctccgaacttcgc-caatttgcttcgaggtccagcgatgctc 98
Db 312 ctccatcgatctctccgaacttcgc-taaactagagagaccaggtaaaagcagtggat 371
QY 99 cacttcgcgcgcgttggtctccacatgacatcatgcatggatggcatttgcaccaatta 158
Db 372 gtggcggtgtgactgcatcctatgctgatgctgctgctgctgctgctgctgctgctgct 431
QY 159 actattgg 166

```

Search completed: November 4, 2000, 13:45:16

```

Db 432 accattgg 439
|| |||||

RESULT 15
X61720
ID X61720 standard; DNA; 573 BP.
XX
AC X61720;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, t457.nt.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX WPI; 1999-189980/16.
XX P-PSDB; Y20023.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 1; Page 165; 275pp; English.
XX
CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 573 BP; 208 A; 43 C; 97 G; 225 T; 0 other;

Query Match 9.4%; Score 36.4; DB 20; Length 573;
Best Local Similarity 51.2%; Pred. No. 0.045;
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 172 cagtattgaaagtgttgagaaagcacacaaagcatattgttgatgtcaccttatgtta 231
Db 140 caaaaattcgtgggttgaaatctcaggcaagagatgatttcattttatcctttttt 199
QY 232 caaatcctcttgatttggtgaaccttggtgcaaaagctggtgctgtgttttacattc 291
Db 200 ttaataatcctaagatagatagataatagtagaataattttctaagggtttgaatttg 259
QY 292 acctagagacatacaagaataaactggaagaacttatccaaagaat 337
Db 260 agttgttataaaataattactttcaaaacggtatagaaaaat 305

```

Sat Nov 4 18:11:21 2000

us-09-300-482-298.rng

Page 10

Job time: 16443 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:33:43 ; Search time 189.35 Seconds  
(without alignments)  
309.911 Million cell updates/sec

Title: US-09-300-482-298

Perfect score: 388

Sequence: 1 ggagaaagaaagaaagatg.....ttaaagcctgggacccccgt 388

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA.\*
- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	58	14.9	1019	5	US-08-714-918-79
C 2	34	8.8	3781	5	US-08-688-988-5
C 3	31.6	8.1	40352	5	US-08-846-111B-15
C 4	31	8.0	1430	1	US-08-276-452A-25
C 5	31	8.0	1430	3	US-08-798-744-25
C 6	30.6	7.9	3150	5	US-08-673-814-5
C 7	30	7.7	2168	5	US-08-749-522-6
C 8	29.8	7.7	2622	2	US-08-766-014-23
C 9	29.4	7.6	5300	2	US-08-766-014-1
C 10	29.4	7.6	6950	4	US-08-841-349-3
C 11	29.4	7.6	8176	4	US-08-841-349-5
C 12	29	7.5	4182	1	US-07-973-257-1
C 13	29	7.5	5181	2	US-08-257-073-10
C 14	29	7.5	8700	3	US-08-392-625-16
C 15	29	7.5	8700	3	US-08-466-961A-16
C 16	28.4	7.3	1347	2	US-08-663-713A-1
C 17	28.4	7.3	1347	5	US-09-014-888-1
C 18	28.4	7.3	2835	2	US-08-750-532-2
C 19	28.4	7.3	4765	2	US-08-750-532-8
C 20	28.4	7.3	10207	2	US-08-920-812-2
C 21	28.4	7.3	10207	2	US-08-920-812-2
C 22	28.4	7.3	10207	2	US-08-921-177-2
C 23	28.4	7.3	10207	2	US-08-362-577C-2
C 24	28.4	7.3	10207	3	US-08-920-828-2
C 25	28.2	7.3	2465	5	US-08-619-812-5
C 26	28	7.2	2403	2	US-08-441-139-17

C 27	28	7.2	2509	4	US-09-014-969-1
C 28	27.8	7.2	5589	1	US-08-465-795-1
C 29	27.8	7.2	5775	2	US-08-306-691B-15
C 30	27.6	7.2	5775	6	PCT-US93-06251-29
C 31	27.6	7.1	1685	1	US-08-105-483-370
C 32	27.6	7.1	1685	2	US-08-709-209-370
C 33	27.6	7.1	1685	2	US-08-458-101-370
C 34	27.6	7.1	8700	4	US-08-645-193B-18
C 35	27.4	7.1	846	4	US-08-743-637B-189
C 36	27.4	7.1	1631	5	US-09-118-319-1
C 37	27.2	7.0	707	6	PCT-US94-06430-25
C 38	27.2	7.0	1359	6	PCT-US94-06430-21
C 39	27.2	7.0	3828	6	PCT-US93-10500-1
C 40	27.2	7.0	5393	4	US-08-591-079-9
C 41	27.2	7.0	7470	5	US-08-417-089-9
C 42	27.2	7.0	7721	5	US-08-772-270A-14
C 43	26.8	6.9	1759	2	US-08-726-525-5
C 44	26.8	6.9	1759	2	US-08-487-942-5
C 45	26.8	6.9	1759	4	US-08-726-036A-5

## ALIGNMENTS

RESULT 1  
US-08-714-918-79/c  
; Sequence 79, Application US/08714918  
; Patent No. 6037123  
; GENERAL INFORMATION:  
; APPLICANT: Benton, Bret  
; APPLICANT: Lee, Ving  
; APPLICANT: Malouin, Francois  
; APPLICANT: Martin, Patrick K.  
; APPLICANT: Schmid, Molly B.  
; APPLICANT: Sun, Dongxu  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL  
; TITLE OF INVENTION: TARGET GENES  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; FILING DATE: September 13, 1996  
; APPLICATION NUMBER: US/08/714,918  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,102  
; FILING DATE: December 22, 1995  
; APPLICATION NUMBER: 60/003,798  
; FILING DATE: September 15, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 222/005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 79:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1019 base pairs  
; TYPE: nucleic acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-714-918-79

```

[illegible]

```

RESULT      2
US-08-688-988-5/c, Application US/08688988B
? Sequence 5, Application US/08688988B
? Patent No. 6096545
? GENERAL INFORMATION:
? APPLICANT: Lefebvre, Daniel D.
? APPLICANT: Malboobi, Mohammad A.
? TITLE OF INVENTION: PHOSPHATE STARV
? FILE REFERENCE: PPL96-03
? CURRENT APPLICATION NUMBER: US/08/6
? CURRENT FILING DATE: 1996-07-31
? NUMBER OF SEQ ID NOS: 48
? SOFTWARE: FASTSEQ for Windows Versi
? SEQ ID NO 5
? LENGTH: 3781
? TYPE: DNA
? ORGANISM: Arabidopsis Thaliana
US-08-688-988-5

```

[illegible]

RESULT 3  
US-08-846-111D-15  
; Sequence 15, Application US/08846111D  
; Patent No. 6017705  
; GENERAL INFORMATION:  
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;  
; APPLICANT: Boon-Falleur, Thierry  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are  
; TITLE OF INVENTION: Members Of The Mage-B Family and Uses Thereof  
; NUMBER OF SEQUENCES: 18

1 CORRESPONDENCE ADDRESS:  
2 ADDRESSEE: Felfe & Lynch  
3 STREET: 805 Third Avenue  
4 CITY: New York City  
5 STATE: New York  
6 ZIP: 10022-7513  
7  
8 COMPUTER READABLE FORM:  
9 MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
10  
11 COMPUTER: IBM  
12 OPERATING SYSTEM: PC-DOS  
13 SOFTWARE: Wordperfect For DOS 6.0b  
14 CURRENT APPLICATION DATA:  
15 APPLICATION NUMBER: US/08/846,111D  
16 FILING DATE: 25-APRIL-1997  
17  
18 CLASSIFICATION: 435  
19 PRIOR APPLICATION DATA:  
20 APPLICATION NUMBER: 08/658,578  
21 FILING DATE: 5-MAY-1996  
22  
23 PRIOR APPLICATION DATA:  
24 APPLICATION NUMBER: 08/403,388  
25 FILING DATE: 14-MARCH-1995  
26  
27 ATTORNEY/AGENT INFORMATION:  
28 NAME: Hanson, No. 6017705man D.  
29 REGISTRATION NUMBER: 30,946  
30 REFERENCE/DOCKET NUMBER: LUD 5444.1  
31  
32 TELECOMMUNICATION INFORMATION:  
33 TELEPHONE: (212) 688-9200  
34 TELEFAX: (212) 838-3884  
35  
36 INFORMATION FOR SEQ ID NO: 15:  
37 SEQUENCE CHARACTERISTICS:  
38 LENGTH: 4052 base pairs  
39 TYPE: nucleic acid  
40 STRANDEDNESS: single  
41 TOPOLOGY: linear  
42  
43 us-08-846-111D-15

	Query Match	8.1%	Score 31.6;	DB 5;	Length 40352;
	Best Local Similarity	54.2%;	Pred. No. 2.9;	Mismatches 0;	Gaps 0;
	Matches 64;	Conservative			
Qy	206	atattggattgtcaccttattgtttacaaaatcctcttgattatgttgaaaccttgacaa	265		
Db	13954	AGATTGGAGGTAGAAACAGCATATTGATGATCCGACCTCTGTGAGCCCTAGAGCTAA	14013		
Qy	266	agctgggtgctcttggttttacatttcacgtgagacatcaaaagataaacttgggaaaga	323		
Db	14014	TGTGGGPGCTTTTGTCTTCATTTTAAACAAAACGCTTTTAAATATTTTAAAGATTAATA	14071		

RESULT 4  
US-08-276-452A-25/c  
; Sequence 25, Application US/08276452A  
; Patent No. 5646029  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Chao-Guang  
; APPLICANT: Mau, Shaio-Lim  
; APPLICANT: Du, He  
; APPLICANT: Gane, Alison M  
; APPLICANT: Basic, Antony  
; APPLICANT: Clarke, Adrienne E  
; TITLE OF INVENTION: Plant Arabinogalactan Protein (AGP) Genes  
; NUMBER OF SEQUENCES: 91  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,452A
; FILING DATE: 18-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Caruthers, Jennie M.
; REGISTRATION NUMBER: 34,464
; REFERENCE/DOCKET NUMBER: 27-91A
; TELEPHONE: (303)499-8080
; TELEFAX: (303)499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1430 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1312
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 41..112
; OTHER INFORMATION: /note= "Derived amino acid sequence
; OTHER INFORMATION: corresponding to the peptide sequence by protein
; OTHER INFORMATION: microsequencing"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 41..112
; OTHER INFORMATION: /note= "Derived amino acids 14-24,
; OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
; OTHER INFORMATION: by protein microsequencing"
; NAME/KEY: misc.feature
; LOCATION: 25..31
; OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
; OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
; OTHER INFORMATION: 26 can be T instead of A"
US-08-276-452A-25

```

```

Query Match      8.0%; Score 31; DB 1; Length 1430;
Best Local Similarity 46.2%; Pred. No. 0.67;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 143 tttgtccccaatttaactattggcgtccagttattgaaagtttgagaaagcacacaaa 202
Db 992 TGTAGTTCACAGTAGGAGTGTGTTTGTGTTGTTGTAATCTCAGAGAAAGCATTTGT 933

QY 203 ggcataattggattgtcaccttatgggtacaaatcctcttgattatgttgaaacctgtgc 262
Db 932 TGTATTGTTTGTGTTGTAATCTCGAGAAACATTTATTGTTGTAGTCTCGGAGA 873

QY 263 aaagctggtcttctggtttacatttcacgtagacatcaaaagataaactggaaga 322
Db 872 AAGCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 813

QY 323 acttatccaaaagaatcaagtcacatggcattctctgtgta 365
Db 812 TGTATTCTCGGAGAAACATTTATTGTTATTATTGTTGTTGTA 770

```

```

RESULT 5
US-08-798-744-25/c
; Sequence 25, Application US/08798744
; Patent No. 5830747
; GENERAL INFORMATION:
; APPLICANT: Chen, Chao-Guang
; APPLICANT: Mau, Shiao-Lim

```

```

; APPLICANT: Du, He
; APPLICANT: Gane, Alison M
; APPLICANT: Bacic, Antony
; APPLICANT: Clarke, Adrienne E
; TITLE OF INVENTION: Plant Arabinogalactan Protein (AGP) Genes
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,744
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/276,452
; FILING DATE: 18-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Caruthers, Jennie M.
; REGISTRATION NUMBER: 34,464
; REFERENCE/DOCKET NUMBER: 27-91A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)499-8080
; TELEFAX: (303)499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1430 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1312
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 41..112
; OTHER INFORMATION: /note= "Derived amino acid sequence
; OTHER INFORMATION: corresponding to the peptide sequence by protein
; OTHER INFORMATION: microsequencing"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 41..112
; OTHER INFORMATION: /note= "Derived amino acids 14-24,
; OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
; OTHER INFORMATION: by protein microsequencing"
; NAME/KEY: misc.feature
; LOCATION: 25..31
; OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
; OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
; OTHER INFORMATION: 26 can be T instead of A"
US-08-798-744-25

```

```

Query Match      8.0%; Score 31; DB 3; Length 1430;
Best Local Similarity 46.2%; Pred. No. 0.67;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

```

QY 143 tttgtccccaatttaactattggcgtccagttattgaaagtttgagaaagcacacaaa 202
Db 992 TGTAGTTCACAGTAGGAGTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 933
QY 203 ggcataattggattgtcaccttatgggtacaaatcctcttgattatgttgaaacctgtgc 262

```

Db 932 TGTATTGTTTGTGTAATTCGGAGAAACAATTATTGTTGTTAGTCTCGGAGA 873  
QY 263 aaaagctgggtcttctgttttacatttcagtagacatcaaaagataactggaaaga 322  
Db 872 AAGCATTGTTGTTATTGTTGTTGTAATTCGGAGAAACAATTATTGTTATTGTTGT 813  
QY 323 acttatcaagaatcaagtcacatgcatgattcctcgtggtga 365  
Db 812 TGTAAATTCGGAGAAACAATTATTGTTATTGTTGTTGTA 770

## RESULT 6

US-08-673-814-5/c  
; Sequence 5, Application US/08673814  
; Patent No. 6086894  
; GENERAL INFORMATION:  
; APPLICANT: Inzana, Thomas J.  
; APPLICANT: Ward, Christine  
; TITLE OF INVENTION: RECOMBINANT VACCINE FOR DISEASES CAUSED  
; TITLE OF INVENTION: BY ENCAPSULATED ORGANISMS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Whitham, Curtis, Whitham & McGinn  
; STREET: 11800 Sunrise Valley Dr., Suite 900  
; CITY: Reston  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 20191  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/673,814.  
; FILING DATE: 27-JUN-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitham, Michael E.  
; REFERENCE/DOCKET NUMBER: VTIP 95-067  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-391-2510  
; TELEFAX: 703-391-9035  
; TELEX: 283072  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3150 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 376..1557  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1586..2740  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2743..3150

## US-08-673-814-5

Query Match 7.9%; Score 30.6; DB 5; Length 3150;  
Best Local Similarity 51.1%; Pred. No. 1.4;  
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 127 acatcatggatggcgatttgcctcccaatttaactattggcgctccagttattgaaagt 186  
Db 1678 AATAACGGATTAGCTTTTAAACCAACTTTTCTGTTTAAACCGGTTTGTAGTTT 1619  
QY 187 tgaagaagacacaaagcgcatattgttgattgtcacctattggttacaaatcctcgtgatt 246

Db 1618 TTCTGTCGACTTCGCGTAATAGTGTGTTCCATATTTTATCCTTATAAATCAATATATT 1559  
QY 247 atgttgaaccccttgqcaaaag 267  
Db 1558 AATAGGCACGAATACGATTAG 1538

## RESULT 7

US-08-749-522-6/c  
; Sequence 6, Application US/08749522  
; Patent No. 6096950  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal  
; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; CITY: Milwaukee  
; STATE: WI  
; COUNTRY: U.S.A.  
; ZIP: 53202-4497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/749,522  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baker, Jean C.  
; REGISTRATION NUMBER: 35,433  
; REFERENCE/DOCKET NUMBER: 670513.90244  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (414) 277-5709  
; TELEFAX: (414) 271-3552  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2168 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-749-522-6

Query Match 7.7%; Score 30; DB 5; Length 2168;  
Best Local Similarity 53.4%; Pred. No. 1.9;  
Matches 63; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 215 ttgtcaccttatggttacaaatcctcttgatttatgttgaaaccttggaagctggtgc 274  
Db 1247 TTAAACATTCCTAATACATCTCTAAATTTATTGATGATTTATATATCAATCAGGTCT 1188  
QY 275 ttctggttttacatttcacgtagacacatcaaaagataactggaaagaaacttatccaa 332  
Db 1187 TTTTGTACTAAATGTTATATATACCGTTAAAAAATGACATGTCAGATCTTATGfAA 1130

## RESULT 8

US-08-766-014-23  
; Sequence 23, Application US/08766014  
; Patent No. 574312  
; GENERAL INFORMATION:  
; APPLICANT: Mamone, Joseph A.  
; APPLICANT: Davis, Maria  
; APPLICANT: Sha, Dan  
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon



Db 3946 ATGGTGGTCATGAAGTCTCTCTCTTTTAATGCGC 3912

## RESULT 11

US-08-841-349-5/c  
; Sequence 5, Application US/08841349B  
; Patent No. 5955594  
; GENERAL INFORMATION:  
; APPLICANT: MISHRA, LOPA  
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.  
; FILE REFERENCE: XX/P04470US0  
; CURRENT APPLICATION NUMBER: US/08/841,349B  
; CURRENT FILING DATE: 1997-04-30  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 8176  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-08-841-349-5

Query Match 7.6%; Score 29.4; DB 4; Length 8176;

Best Local Similarity 56.8%; Pred. No. 6.4; Mismatches 0; Gaps 0;  
Matches 54; Conservative 0; Indels 41; Indels 0; Gaps 0;

Qy 73 tggcttcgagctcagcgatgctccacttcgcccgcgattggctccacatggacatca 132

Db 4006 TCGCTCACCAGTCCTCGGCCAGTCTCCACACACGATGATCTCTCCTGGTGGCATCC 3947

Qy 133 tggatggcatttggcccccaatttaactattggc 167

Db 3946 ATGGTGGTCATGAAGTCTCTCTCTTTTAATGCGC 3912

## RESULT 12

US-07-973-257-1/c  
; Sequence 1, Application US/07973257  
; Patent No. 5378820  
; GENERAL INFORMATION:  
; APPLICANT: Keeler, Jr. Calvin L.  
; APPLICANT: Dolms, John E.  
; TITLE OF INVENTION: Gene Encoding Cytadhesin  
; TITLE OF INVENTION: Protein of Mycoplasma Gallisepticum and Its Use  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Connolly and Hutz  
; STREET: 1220 Market Street  
; CITY: Wilmington  
; STATE: Delaware  
; COUNTRY: U.S.A.  
; ZIP: 19899  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM/PC or Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/973,257  
; FILING DATE: 19921109  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: No. 5378820e  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4182 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double stranded  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: No  
; ANTI-SENSE: No  
; ORIGINAL SOURCE:  
; ORGANISM: Mycoplasma Gallisepticum

; STRAIN: S6  
US-07-973-257-1

Query Match 7.5%; Score 29; DB 1; Length 4182;  
Best Local Similarity 57.0%; Pred. No. 5.9;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 206 atatttgattgcacattgattacaaatcctcttcaattatgtgaaccttggcaaa 265

Db 4102 ATCTTAATAATATACCTTAATATTCAATTTCTTAGTGTGCTGAGAGCTTTGGTTT 4043

Qy 266 agctggctctctggttttacatttcacgtaga 298

Db 4042 AGCTGGTCTGTTGGTTTTCCTCCTGGAGACAGA 4010

## RESULT 13

US-08-257-073-10  
; Sequence 10, Application US/08257073  
; Patent No. 5766597  
; GENERAL INFORMATION:  
; APPLICANT: Paoletti, Enzo  
; APPLICANT: de Taisne, Charles  
; APPLICANT: Tine, John A.  
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue, 25th Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/257,073  
; FILING DATE: 09-JUN-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/075,783  
; FILING DATE: 11-JUN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/852,305  
; FILING DATE: 18-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,183  
; FILING DATE: 20-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2570  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; TELEX: 425066 CURTMS  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5181 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-257-073-10

Query Match 7.5%; Score 29; DB 2; Length 5181;  
Best Local Similarity 53.0%; Pred. No. 6.7;  
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;



Search completed: November 4, 2000, 13:33:58  
Job time: 16674 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:10 ; Search time 4352.3 Seconds  
(without alignments)  
551.186 Million cell updates/sec

Title: US-09-300-482-298

Perfect score: 388

Sequence: 1 ggagagaagaagaaagatg.....ttaagcctggagaccccgct 388

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estpl1.\*  
71: em\_estpl2.\*  
72: em\_estpl3.\*  
73: em\_estpl4.\*  
74: em\_estpl5.\*  
75: em\_estro1.\*  
76: em\_estro2.\*  
77: em\_estro3.\*  
78: em\_estro4.\*  
79: em\_estro5.\*  
80: em\_estro6.\*  
81: em\_estro7.\*  
82: em\_estro8.\*  
83: em\_estro9.\*  
84: em\_estro10.\*  
85: em\_estro11.\*  
86: em\_estro12.\*  
87: em\_estro13.\*  
88: gb\_gss1.\*  
89: gb\_gss2.\*  
90: gb\_gss3.\*  
91: gb\_gss4.\*  
92: em\_gss1.\*  
93: em\_gss2.\*  
94: em\_gss3.\*  
95: em\_gss4.\*  
96: gb\_gss5.\*  
97: gb\_gss6.\*  
98: gb\_gss7.\*  
99: gb\_gss8.\*  
100: gb\_gss9.\*  
101: em\_gss5.\*  
102: em\_gss6.\*  
103: em\_gss7.\*  
104: em\_gss8.\*  
105: em\_gss9.\*  
106: em\_gss10.\*  
107: em\_gss11.\*  
108: gb\_gss10.\*  
109: gb\_gss11.\*  
110: em\_gss12.\*  
111: gb\_gss12.\*  
112: gb\_gss13.\*  
113: gb\_gss14.\*  
114: gb\_gss15.\*  
115: gb\_gss16.\*  
116: gb\_gss17.\*

LOCUS  
DEFINITION  
Sa34c12.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-cl004-1199 5' similar to TR:014105 014105 RIBULOSE-PHOSPHATE  
3-EPIMERASE. ; , mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AI437973 418 bp mRNA EST 18-APR-2000  
AI437973.1 GI:4289768  
EST  
soybean.  
Glycine max

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE  
AUTHORS

1 (bases 1 to 418)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna  
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers  
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
R., Waterston, R. and Wilson, R.

TITLE  
JOURNAL  
COMMENT

Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com  
Insert Length: 945 Std Error: 0.00  
Seq Primer: -40RP from Gibco  
High quality sequence stop: 414  
POLYA=No.

Location/Qualifiers  
1. 418  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl004-1199"  
/clone\_lib="Gm-cl004"  
/tissue\_type="root"  
/lab\_host="XL10-Gold"  
/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:  
XhoI; Root cDNA. The mRNA was isolated from entire roots  
of 8 day old 'Williams' seedlings which were propagated on  
paper towels with distilled water. Stratagene's cDNA  
Synthesis Kit (catalog #200401) was used to synthesize the  
cDNA. First-strand synthesis was performed with 5-methyl  
dCTP, hence the ligated cDNA is hemimethylated.  
Stratagene's first-strand synthesis primer was used  
[GAGAGAGAGAGAGAGAGAGACTGCTCGAG(T)-18]. After  
second-strand synthesis, the cDNA ends were 'polished'  
with clone Pru DNA polymerase, ligated to EcoRI adapters,  
and phosphorylated. The XhoI site within the first-strand  
synthesis primer was restricted by digestion with XhoI;  
all XhoI sites in the cDNA would be protected by their  
hemimethylated status. The cDNA constructs were  
size-fractionated with a 500bp cutoff, using GibcoBRL Life  
Technologies' cDNA Size fractionation column. The column  
eluent was then ligated into Stratagene's pBluescript II  
XR Predigested vector (pBluescript II SK(+)) that had been  
digested with EcoRI and XhoI, and phosphorylated. Both  
the white and blue colonies appear to contain recombinant  
plasmids with cDNA inserts. Blue colonies 9n-15) have been  
sequenced, and possess putative cDNA inserts. This library  
was constructed by Dr. Paul Keim & Virginia H. Coryell,  
Department of Biology, Box 5640, Northern Arizona  
University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.  
Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:  
520-523-7500, email: paul.keim@naui.edu,

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	365.4	94.2	418	10	AI437973 sa34c12.y
2	320.8	82.7	403	18	AV412302
3	310.4	80.0	405	10	AI442690
4	307.8	79.3	641	23	AW682881
5	296.8	76.5	633	34	BE325053
6	285.4	73.6	736	21	AW348508
7	281.2	72.5	399	18	AV425192
8	281.2	72.5	448	14	AL371120
9	244.6	62.6	494	35	BE404347
10	243	62.6	494	35	BE404347
11	238.2	61.4	526	36	BE494864
12	237.4	61.2	407	15	AO069503
13	226.8	58.5	483	19	AV554048
14	226	58.2	497	11	AI490023
15	226	58.2	552	12	AI772355
16	226	58.2	593	19	AW031337
17	226	58.2	630	20	AW223738
18	226	58.2	653	13	AI896020
19	225	58.0	483	35	BE353381
20	225	58.0	552	11	AI485294
21	224.4	57.8	520	12	AI771373
22	222.4	57.3	569	19	AV553860
23	222	57.2	451	36	D41947
24	221.8	57.2	508	11	AI485741
25	220.6	56.9	450	13	AI898953
26	220	56.7	378	9	AI164978
27	212.4	54.7	434	9	AI164426
28	211.8	54.6	522	9	AI162283
29	207.4	53.5	570	9	AI161491
30	203	52.3	380	35	BE358283
31	201.6	52.0	348	35	BE404517
32	198.4	51.1	457	36	D41569
33	197.6	50.9	352	36	D24955
34	187.8	48.4	725	19	AW010209
35	174	44.8	272	9	AI164995
36	172.4	44.4	638	21	AW350186
37	158.4	40.8	613	5	AW556420
38	152	39.2	268	9	AI165092
39	150.2	38.7	441	9	AI161745
40	148.8	38.4	209	9	AI165543
41	147	37.9	544	19	AV545174
42	128.4	33.1	232	36	D41516
43	116.8	30.1	749	33	BE034126
44	101.8	26.2	170	20	AW224026
45	101	26.0	465	36	C88722

## ALIGNMENTS

BASE COUNT 111 a 93 c 101 g 113 t  
 ORIGIN virginia.coryell@nau.edu"

Query Match 94.2%; Score 365.4; DB 10; Length 418;  
 Best Local Similarity 97.1%; Pred. No. 3.3e-99;  
 Matches 372; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 6 aagaagaagaatggaatgacaccgaaatagctcttccttcgacttc 65  
 DB 4 AGGAAGGAAGATGGAGTACACCGAAAAATGCTCTTCGATGCTCTCCGACTTC 63

QY 66 gcaaatgttgcctccgagctcagcgatgctccacttcgcccacattgctccacatg 125  
 DB 64 GCCAATTTGGCTTCGAGGCTCAGGCGATGCTCCACTTCGGCGCGGATGGCTCCACATG 123

QY 126 gacatcgtggtggcatttggctcccaatttaactattggcgctccagttattgaaagt 185  
 DB 124 GACATCATGGATGGGCAATTTGTCCCAATTTAACTATTGGCGCTCCAGTTATTGAAAGT 183

QY 186 ttgagaagcacacaaaggcatatttgattgtcaccttattggtttacaaatcctttgat 245  
 DB 184 TTGAGAAGCACACAAAGGATATTGGATGTGACCTTATGTTTACAAATCCTCTTGAT 243

QY 246 tatgtgaaccttggcacaagctgtgcttctggtttacatttcacgttagagacatca 305  
 DB 244 TATGTTGAGCCCTTGGCAAAAGCTGGTCTTCTGGTTTACATTTTATGTAGAGACATCA 303

QY 306 aagataactggaagaacttaccaaagaatcaagtcacatgcatgcttctgtgtga 365  
 DB 304 AAAGTAACTGGGAAGAACATTATCCAAAGAAATCAAGTCACATGGCATGACTCCTGTGTGA 363

QY 366 gcattaaagcctggagcccccgt 388  
 DB 364 GCATTAAGCTGGGACCCCAT 386

## RESULT 2

AV412302 403 bp mRNA EST 23-MAY-2000  
 LOCUS AV412302 Lotus japonicus young plants (two-week old) Lotus  
 DEFINITION japonicus cDNA clone MW217c10\_r 5', mRNA sequence.

ACCESSION AV412302  
 VERSION AV412302  
 KEYWORDS EST.  
 SOURCE AV412302.1 GI:7741466  
 ORGANISM Lotus japonicus.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.

REFERENCE 1 (bases 1 to 403)  
 ASAMIZU, E., NAKAMURA, Y., SATO, S. and TABATA, S.  
 Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus  
 JOURNAL DNA Res 7 (2), 127-130 (2000)  
 MEDLINE 20277479  
 COMMENT Contact: Yasukazu Nakamura  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 Location/Qualifiers  
 ..403  
 /organism="Lotus japonicus"  
 /db\_xref="taxon:34305"  
 /clone="MW217c10\_r"  
 /clone\_lib="Lotus japonicus young plants (two-week old)"  
 /dev\_stage="young plants (two-week old)"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI; Isolate-Miyakojima MG-20"

BASE COUNT 113 a 93 c 91 g 106 t  
 ORIGIN

Query Match 82.7%; Score 320.8; DB 18; Length 403;  
 Best Local Similarity 91.4%; Pred. No. 7.9e-86;  
 Matches 340; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 4 gaaagaagaagaatggaatgacaccgaaatagctcttccttcgacttccttcgact 63  
 DB 32 GTAACAAACAAACGATGGAGTGACACCCAAATTTGCTCGTGTGCTCTCATCGGACT 91

QY 64 tgcacaatttggcttcgagagctcagcgcatgctccacttcggcgccgattggctccaca 123  
 DB 92 TCGCCAATTTGGCTTCGAGGCTCAACGCATGCTTGATTATGGCGCGGATGGCTCCACA 151

QY 124 tggacatcgtatggcgcatatttgcctcccaatttaactattggcgctccagttattgaaa 183  
 DB 152 TGGACATCATGATGGGCACTTTGTCCCAATTTAACTATTGGCACTTCGGGTCAATTGAAA 211

QY 184 gttgagaagcacacaaaggcatatttggattgtcaccttattggttacaatcctcttg 243  
 DB 212 GTTTGAGAAGCACACAAAGGCATATCTGGATTGTACCTTATGTTTACAAATCCTCTTG 271

QY 244 attatgtgaaccttggcacaagctgtgcttcttggttttacatttcacgttagagacat 303  
 DB 272 ATATGTTGAGCCCTTGGGAAAGCAGGTGCTTCTGGTTTACATTTTATGTAGAGCAT 331

QY 304 caaagataactggaagaacttattccaaagaatcaagtcacatgcatgattcctctgtgtg 363  
 DB 332 CAAAAGATAACTGGAAAGAAGTTATCCAAAGAATTAACTACAGGCGCATGAAGCCTGGTG 391

QY 364 tagcattaaagc 375  
 DB 392 TAGCATTTAAAGC 403

## RESULT 3

AI442690 405 bp mRNA EST 18-APR-2000  
 LOCUS sa85b04.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 DEFINITION Gm-cl004-6056 5' similar to TR:014105 014105 RIBULOSE-PHOSPHATE 3-EPIMERASE. ;, mRNA sequence.

ACCESSION AI442690  
 VERSION AI442690  
 KEYWORDS AI442690.1 GI:4297942  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 405)  
 SHOEMAKER, R., KEIM, P., VODKIN, L., ERPELDING, J., CORYELL, V., KHANNA, A., BOLLA, B., MARRA, M., HILLIER, L., KUCABA, T., MARTIN, J., BECK, C., WYLIE, T., UNDERWOOD, K., STEPTOE, M., THEISING, B., ALLEN, M., BOWERS, Y., PERSON, B., SWALLER, T., GIBBONS, M., PAPE, D., HARVEY, N., SCHURK, R., RITTER, E., KOHN, S., SHIN, T., JACKSON, Y., CARDENAS, M., MCCANN, R., WATERSTON, R. and WILSON, R.  
 Public Soybean EST Project  
 Unpublished (1999)  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

TITLE This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 JOURNAL Insert Length: 923 Std Error: 0.00  
 COMMENT Seq primer: -40RP from Gibco  
 High quality sequence stop: 395

POLYA=NO. Location/Qualifiers

1. .405  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl004-6056"  
 /clone\_lib="Gm-cl004"  
 /tissue\_type="root"  
 /lab\_host="XL10-Gold"  
 /note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2: XhoI; Root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratiagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratiagene's first-strand synthesis primer was used [GAGAGAGAGAGAGAGAGACTAGTCTCGAG(T)-18]. After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratiagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Corryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Corryell), Fax: 520-523-7500, email: paul.keim@naui.edu, virginia.corryell@naui.edu"

BASE COUNT 111 a 86 c 96 g 112 t

ORIGIN

Query Match 80.0%; Score 310.4; DB 10; Length 405;  
 Best Local Similarity 91.6%; Pred. No. 1.1e-82;  
 Matches 348; Conservative 0; Mismatches 11; Indels 21; Gaps 1;

Qy 9 aaagaaagatggaatgacacgaaatagctccttcgatgctcttcgcactgcgc 68  
 |||||  
 Db 1 AAAGGAAGGATGGAGTGACACCGAAATGCTCTCTTCGATGCTCTTCGACTTCGCC 60  
 |||||  
 Qy 69 aatttgcttcgcaggctcagcgatgctccacttcggtcgcgagtggttcacatggac 128  
 |||||  
 Db 61 AATTGGCTTCGAGGCTCAGCGCAT-----GCTCCACATGGAC 99  
 |||||  
 Qy 129 atcatgtaggggattttgcccccaatttaactatttgccctccacttattgaaagttag 188  
 |||||  
 Db 100 ATCATGATGGGATTTTGTCCCAATTTAACTATTGGGCGCTCCAGTTATTGAAGTTTG 159  
 |||||  
 Qy 189 agaaagcacaaaggcatatttggttggttcacattatggttacaaatcctcttgattat 248  
 |||||  
 Db 160 AGAAGCACAAAGGATATTGGATTGTTCACCTTATGTTTACAAATCCTCTTGATTAT 219  
 |||||  
 Qy 249 gttgaacccctggcaaaagctggtgctctggttttacattttcaacgtagagacatacaaa 308  
 |||||  
 Db 220 GTTGAGCCCTTGGCAAAAGCTGTGCTTCTGGTTTACATTTTCATGTAGAGACATCAAA 279  
 |||||  
 Qy 309 gataactggaaagaactatccaaagaatacaagtcaacatgcatgattcctggtgtagca 368  
 |||||  
 Db 280 GATAACTGGGAGAACATTATCAAGAAATCAAGTCACATGGCATGGATGCTCTGTGTAGCA 339  
 |||||  
 Qy 369 ttaagcctgggagcccccg 388  
 |||||  
 Db 340 TTAAGGCTGGGAGCCCAT 359  
 |||||

RESULT 4 Location/Qualifiers

AW682881 641 bp mRNA EST 15-JUN-2000  
 LOCUS NF001C10L1F1081 Developing leaf Medicago truncatula cDNA clone  
 DEFINITION NF001C10L1F 5', mRNA sequence.  
 AW682881  
 ACCESSION AW682881.1 GI:7557597  
 VERSION  
 KEYWORDS barrel medic.  
 SOURCE Medicago truncatula  
 ORGANISM  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;  
 Fabales; Fabaceae; Papilionoideae; Medicago.  
 REFERENCE 1 (bases 1 to 641)  
 AUTHORS Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,  
 Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.  
 TITLE Expressed Sequence tags from the Samuel Roberts Noble Foundation  
 JOURNAL Medicago truncatula leaf library  
 COMMENT Unpublished (2000)  
 CONTACT: May GD  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7391  
 Fax: 580 221 7380  
 Email: gdmay@noble.org  
 Insert Length: 641 Std Error: 0.00  
 Plate: 001 row: C column: 10  
 Seq primer: TCACACAGGAACAGCTATGAC.  
 Location/Qualifiers  
 1. .641  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone="NF001C10L1F"  
 /clone\_lib="Developing leaf"  
 /tissue\_type="leaf"  
 /dev\_stage="Pooled developmental"  
 /note="Vector: Lambda Zap; Contains a mixture of very  
 young, developing, mature and senescing leaves."

BASE COUNT 201 a 130 c 139 g 171 t

ORIGIN

Query Match 79.3%; Score 307.8; DB 23; Length 641;  
 Best Local Similarity 87.7%; Pred. No. 7e-82;  
 Matches 336; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 6 aagaaagaaagatgggaatgacacccgaaatagctccttcgatgctcttcgcacttc 65  
 |||||  
 Db 126 AAAAGAGAAGATGGAGTGACACCGAAATAGCTCTTCGATGCTATCATCAGATTTT 185  
 |||||  
 Qy 66 gccaatgtgcttcgcaggctcagcgcatgctccacttcggtccgagattggtccacatg 125  
 |||||  
 Db 186 GCTAATTTGGCTCCGAAGCTCATGCTATGATCAATTAATACGCGCTGATTTGGCTTCATG 245  
 |||||  
 Qy 126 gacatcatgtagtggaattttgtcccaatttaactatttgccgtccagttattgaaagt 185  
 |||||  
 Db 246 GATATCATGATGGGCAATTTTGTCCCTTAATTAATGAGGCGCTCCAAATCATTGAAAGT 305  
 |||||  
 Qy 186 ttgagaagcacacaaaggcatatttggttggttcacattatggttacaaatcctcttgat 245  
 |||||  
 Db 306 TTGAGAAGCACACAGAGGCATATCTGGACTGTCCCTGATGTTTACAAATCCACTTGTAT 365  
 |||||  
 Qy 246 tatgtgaacccttggaacaaagctggtgctcttggtttttacatttcacgtagagacatca 305  
 |||||  
 Db 366 TATGTTGAACCTTGGGAAAAGCTGGTCTCTGTTTACATTTTACATTTATATAGACATCA 425  
 |||||  
 Qy 306 aaagataactggaagaacttatccaaagaatacaagtcaacatgcatgattcctggtgtga 365  
 |||||  
 Db 426 AAAGACAACCTGGAAGAAGCTTATCCAAAATATTAAGTCACACGCGCATGAGCGCTGGTGA 485  
 |||||

QY 366 gattataaagcctgggacccccgt 388  
 Db 486 TCGATAAAGCCTGGACATCCGT 508

RESULT 5  
 BE325053 633 bp mRNA EST 14-JUL-2000  
 LOCUS NF119G01S1F1006 Developing stem Medicago truncatula cDNA clone  
 DEFINITION NF119G01S1 5', mRNA sequence.  
 ACCESSION BE325053  
 VERSION BE325053.1 GI:9198914  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 Fabales; Fabaceae; Papilionoideae; Medicago.  
 REFERENCE 1 (bases 1 to 633)  
 AUTHORS He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell  
 ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon  
 ,R.A.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 JOURNAL Medicago truncatula stem library  
 COMMENT Unpublished (2000)  
 CONTACT: Dixon RA  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7302  
 Fax: 580 221 7380  
 Email: radixon@noble.org  
 MEDICAGO Genome Initiative accession: MGI:S:15815  
 Insert Length: 633 Std Error: 0.00  
 Plate: 119 row: 6 column: 01  
 Seq primer: TCACAGGAAACAGCTATGAC.  
 Location/Qualifiers  
 1. 633  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone="NF119G01S1"  
 /clone\_lib="Developing stem"  
 /tissue\_type="stem"  
 /dev\_stage="Pooled developmental"  
 /note="Vector: Lambda Zap; Contains a mixture of  
 internodal stem segments"  
 BASE COUNT 194 a 131 c 138 g 170 t  
 ORIGIN

Query Match 76.5%; Score 296.8; DB 34; Length 633;  
 Best Local Similarity 87.5%; Pred. No. 1.4e-78;  
 Matches 336; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

QY 6 aagaagaagaatgggaatgacacgaaataagctcttcgattgctcttccttcgacttc 65  
 Db 104 AAAAGAAGAAGATGGGAGTGACACCCGAAATAGCTCTTCGATGTCATCATCAGATTIT 163

QY 66 gccaatttgctccagagctcagcagctcctcacttcgagcgcgattggctccacatg 125  
 Db 164 GCTAATTTGGCTCCGAGCTCATCGTATGATCAATTAACCTGGCGCTGATGGCTTCACATG 223

QY 126 gacatcatgattgggcatattgtcccaatttaactattggcctcagattgaaagt 185  
 Db 224 GATATCATGATGGCATTTTGTCCCTTAATTAACCTGGCGCTCCATCATGAAAT 283

QY 186 ttgagaagacacaaaggcatttggattgtcacttattggttacaattcctcttggat 245  
 Db 284 TTGAGAAGACACAGAGGATATCTGGAGTGTCACTGATGGTTACAAATCCACTTGTAT 343

QY 246 tatgttgaaaccttggcaaaagctgtgctcttggttttacctttacatttcacgtagacatca 305  
 Db 344 TATGTTGAACCTTGGGAAAAGCTGGTGTCTTCGGTTTACATTTTCATATAGAGACATCA 403

QY 306 aaagataactggaaagaacttatcc-aaagaatcaagtcacatgcatgattcctgtgt 364  
 Db 404 AAGACAACTGGAAGAACTTATCCAAATATTAAGTCACACGGCATGAGGCTGTGT 463

QY 365 agcattaaagcctgggacccccgt 388  
 Db 464 ATCGATAAAGCCTGGACATCCGT 487

RESULT 6  
 AW348508/c 736 bp mRNA EST 01-FEB-2000  
 LOCUS GM210002B12C3R Gm-r1021 Glycine max cDNA 3', mRNA sequence.  
 DEFINITION AW348508  
 ACCESSION AW348508  
 VERSION AW348508.1 GI:6846218  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 736)  
 AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,  
 Erpelting,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.  
 TITLE A Functional Genomics Program for Soybean (NSF 9872565)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Vodkin, L.O., PI, A Functional Genomics Program for  
 Soybean (NSF 9872565)  
 Lewin, H. A., Director, Keck Center for Comparative and Functional  
 Genomics  
 University of Illinois  
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
 Tel: (217) 244-6147  
 Fax: (217) 333-4582  
 Email: l-vodkin@uiuc.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134. For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888)919-3324 or (314)  
 427-3324 or contact:clones@genomesystems.com or info@genome  
 systems.com web site:www.genomesystems.com  
 Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

Location/Qualifiers  
 1. 736  
 /organism="Glycine max"  
 /cultivar="Williams"  
 /db\_xref="taxon:3847"  
 /clone\_lib="Gm-r1021"  
 /tissue\_type="root"  
 /lab\_host="XL10-Gold"  
 /note="Vector: pBluescript II XR; Site:1: EcoRI; Site:2:  
 XhoI; Library Gm-r1021 is a sequence-driven, reracked set  
 of the original library Gm-cl004 which was prepared from  
 root cDNA. The mRNA was isolated from entire roots of 8  
 day old 'Williams' seedlings which were propagated on  
 paper towels with distilled water. StrataGene's cDNA  
 Synthesis Kit (catalog #200401) was used to synthesize  
 the cDNA. The Gm-cl004 library was constructed by Dr.  
 Paul Keim & Virginia H. Coryell, Department of Biology,  
 Box5640, Northern Arizona University, Flagstaff, AZ 86011  
 , email: paul.keim@uau.edu, virginia.coryell@uau.edu. The  
 contig analysis to select unique genes was performed by the  
 the Laboratory of Ernest Retzel, Computational Biology  
 Centers, University of Minnesota,  
 http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html  
 . Reracking was performed by Genome Systems, St. Louis,  
 http://www.genomesystems.com, and sequencing by the Keck  
 Center for Comparative and Functional Genomics,  
 University of Illinois,  
 http://www.life.uiuc.edu/biotech/keck.html."  
 197 a 164 c 131 g 215 t 29 others

BASE COUNT  
 ORIGIN





```

/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE041_B06_C11"
/clone_lib="Wheat etiolated seedling root cDNA library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized,
germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Roots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared, a cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

```

```

BASE COUNT      113 a   127 c   143 g   111 t
ORIGIN

```

```

Query Match      62.6%; Score 243; DB 35; Length 494;
Best Local Similarity 79.3%; Pred. No. 1.9e-62;
Matches 288; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 26 gacacgaaatagctccttcgatgctcttcgacttcgccaattggcttcgagggc 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 53 GCGCGGAGATAGCGCGTCCATGCTGCTCGGACTTCGCCAACCTCGCTCGGAGGC 112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 86 tcagcgaatgcctccacttcggcgccgattggctccacatggaacatcatgatggcattt 145
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 113 CGAGCGATGTCGGCTCGCGCGCGACTGGCTCCACATGCGACGTCATGATGGGCACT 172
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 146 tgcctcaatttaactattgctcctcagttattgaaattgagaaagcacacaaagggc 205
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 173 TGTTCCTAATTACTATGAGCTCCAGTGTGAGAGCTTGAGAGGACACACAAAGGC 232
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 206 atatttgattgacaccttgggtacaaatcctcttgattattgtgaaccttggcaaa 265
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233 ATATTGAGCTGCCATCTCATGGTGCACAAATCCTTCAGATTATGTAGAGGCATTTGGAAA 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 266 agctggtctctctgttttaattcaatgtagacatcaaaagataaactggaagaaact 325
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 293 AGCTGGTCCCTCAGGATTCACGTTCCATATAGAAAGTAACAGGGGTAACCTGGAAGAGCT 352
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 326 tatcaagaatcaagtcacatggtcattcctcgtgttagcattaaagcctgggacccc 385
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 353 CATCCAAAGCATTAAGCAAGGGCATCGACCTGGTGTATCATTTGAAGCCTGGTACTCC 412
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 386 cgt 388
DB 413 TGT 415

```

```

RESULT 11
BE494864      526 bp      mRNA      02-AUG-2000
LOCUS        WHE1258_D10_H20S Secale cereale anther cDNA library Secale cereale
DEFINITION   cDNA clone WHE1258_D10_H20, mRNA sequence.
ACCESSION   BE494864
VERSION     BE494864.1 GI:9661373
KEYWORDS     EST.
SOURCE       rye.
ORGANISM     Secale cereale
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Secale.
REFERENCE   1 (bases 1 to 526)
AUTHORS     Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton
            R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R.,
            Miller,R., Rausch,C.J., Ross,K., Seaton,C.I. and Tong,J.C.

```

# TITLE JOURNAL COMMENT

The structure and function of the expressed portion of the wheat  
genomes - Anther cDNA library from rye  
Unpublished (2000)  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@pw.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stragatene SK primer.

## FEATURES source

```

1..526
Location/Qualifiers
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE1258_D10_H20"
/clone_lib="Secale cereale anther cDNA library"
/tissue_type="Anther"
/dev_stage="Adult plant before anthesis"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Anthers were harvested and pooled from early
meiosis to late meiosis. The tissue, total RNA, and
poly(A) RNA were prepared (Butler, Ross and Gustafson) at
University of Missouri, Columbia. A cDNA library was
made, and the cDNA clones were in vivo excised to give
pBluescript phagemids in the TJ Close lab (Choi, Close,
Fenton) at the University of California, Riverside.
Plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson lab (all other authors)."

```

```

BASE COUNT      123 a   130 c   157 g   116 t
ORIGIN

```

```

Query Match      61.4%; Score 238.2; DB 36; Length 526;
Best Local Similarity 78.5%; Pred. No. 5.3e-61;
Matches 285; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 26 gacacgaaatagctccttcgatgctcttcgacttcgccaattggcttcgagggc 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 12 GCGCGGAGATAGCGCGTCCATGCTGCTCGGACTTCGCCAACCTCGCGCAGAGGC 71
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 86 tcagcgaatgcctccacttcggcgccgattggctccacatggaacatcatgatggcattt 145
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72 CGAGCGATGTCGGCTCGCGCGCGACTGGCTCCACATGAGAGCTTGAGAGGCATTTGGGAA 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 146 tgcctcaatttaactattgctcctcagttattgaaattgagaaagcacacaaagggc 205
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132 CGTTCCTAATTACTATGAGCTCCAGTGTGAGAGCTTGAGAGGCATTTGGGAA 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 206 atatttgattgacaccttgggtacaaatcctcttgattattgtgaaccttggcaaa 265
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 ATATTGAGCTGCCATCTCATGGTGCACAAATCCTTCCTGATTATGTAGAGGCATTTGGGAA 251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 266 agctggtctctctgttttaattcaatttcacatggaacatcaaaagataaactggaagaaact 325
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 AGCTGGCGCTTCAGGATTCACATTCCTATAGAAAGTAGCAGGGAACACTGGAAAGAGCT 311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 326 tatcaagaatcaagtcacatggtcattcctcgtgttagcattaaagcctgggacccc 385
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 CATCCAAAGCATTAAGCAAGGGCATCGCGCTGGTGTATCATTTGAAGCCTGGTACTCC 371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 386 cgt 388
DB 372 TGT 374

```

```

RESULT 12
AU069503

```



LOCUS AV069503 407 bp mRNA EST 07-JUN-1999  
DEFINITION AV069503 Rice callus Oryza sativa cDNA clone C62255\_11A, mRNA  
sequence.  
ACCESSION AV069503  
VERSION AV069503.1 GI:5004354  
KEYWORDS EST.  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
AUTHORS Yamamoto,K. and Sasaki,T.  
TITLE Rice cDNA from callus (1998)  
JOURNAL Unpublished (1998)  
COMMENT Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai,Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7458  
Email: tsasaki@br.affrc.go.jp  
PROJECT = "RGP".  
FEATURES  
source  
1. .407  
Location/Qualifiers  
/organism="Oryza sativa"  
/strain="cultivar Nipponbare, sub\_species Japonica"  
/db\_xref="taxon:4530"  
/clone="C62255\_11A"  
/clone\_lib="Rice callus"  
/note="Vector: pBluescript II SK+; Site\_1: SalI; Site\_2:  
NotI; cDNA prepared from rice callus mRNAs by using  
oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid."  
99 a 106 c 116 g 86 t

Query Match 61.28; Score 237.4; DB 15; Length 407;  
Best Local Similarity 78.8%; Pred. No. 8.7e-61;  
Matches 283; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
Qy 26 gacaccgaataagctccttcgactgctctccgacttcgcaattgcttcccaaggc 85  
Db 49 GCGGCGAAGATAGCGCGTCGATGCTCTCGTCGAGCTTCGCCAACCTCCCGGAGGC 108  
Qy 86 tcagcgatgctccacttcgcgccgattggctccacatgacatcatgagtgatg 145  
Db 109 CGACCGCATGGTCCGCTCGCGCGCGACTGGCTCCACATGCACATGACGGCACCT 168  
Qy 146 tgcaccaatttaactattggcgtccagttattgaagttagaagttgagaaagcacacaaggc 205  
Db 169 TGTTCCTAATCTACTATTGAGCTCCAGTGATTCAGAGCTTGAGGAAGCACACCAAGGC 228  
Qy 206 atattggattgcacattatggttaacaaatcctcttgattgattgaaaccttgcaaa 265  
Db 229 ATATTGGACTGCCATCTTATGTTGACCAATCTTCGGATTATGTAGAACCATTAGCAAA 288  
Qy 266 agctgggtcttctggtttacattcaacgttagagacatcaaaagataaactggaaagact 325  
Db 289 AGCTGGTGCCTCAGGTTTCATTCATATAGAGATATCCAGAGACAAATTTGGCAAGACT 348  
Qy 326 tatccaaagaatacaagtcacatggcattccttggttagcattaaagcctgggacc 384  
Db 349 CATCCAAGTATCAACCAAGGATGCGACCGGGTGTATCATTTGAGGCCAGGCACCTC 407

LOCUS AV554048 483 bp mRNA EST 23-JUL-2000  
DEFINITION AV554048 Arabidopsis thaliana roots Columbia Arabidopsis thaliana  
cDNA clone RZ79e04R 5', mRNA sequence.  
ACCESSION AV554048  
VERSION AV554048.1 GI:8725461  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsis.  
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
of 12,028 non-redundant expressed sequence tags from normalized and  
size-selected cDNA libraries  
JOURNAL DNA Res. 7, 175-180 (2000)  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
FEATURES  
source  
1. .483  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="RZ79e04R"  
/clone\_lib="Arabidopsis thaliana roots Columbia"  
/tissue\_type="roots"  
/note="Vector: pBluescriptII SK+; Site\_1: EcoRI; Site\_2:  
XhoI"  
121 a 108 c 132 g 122 t

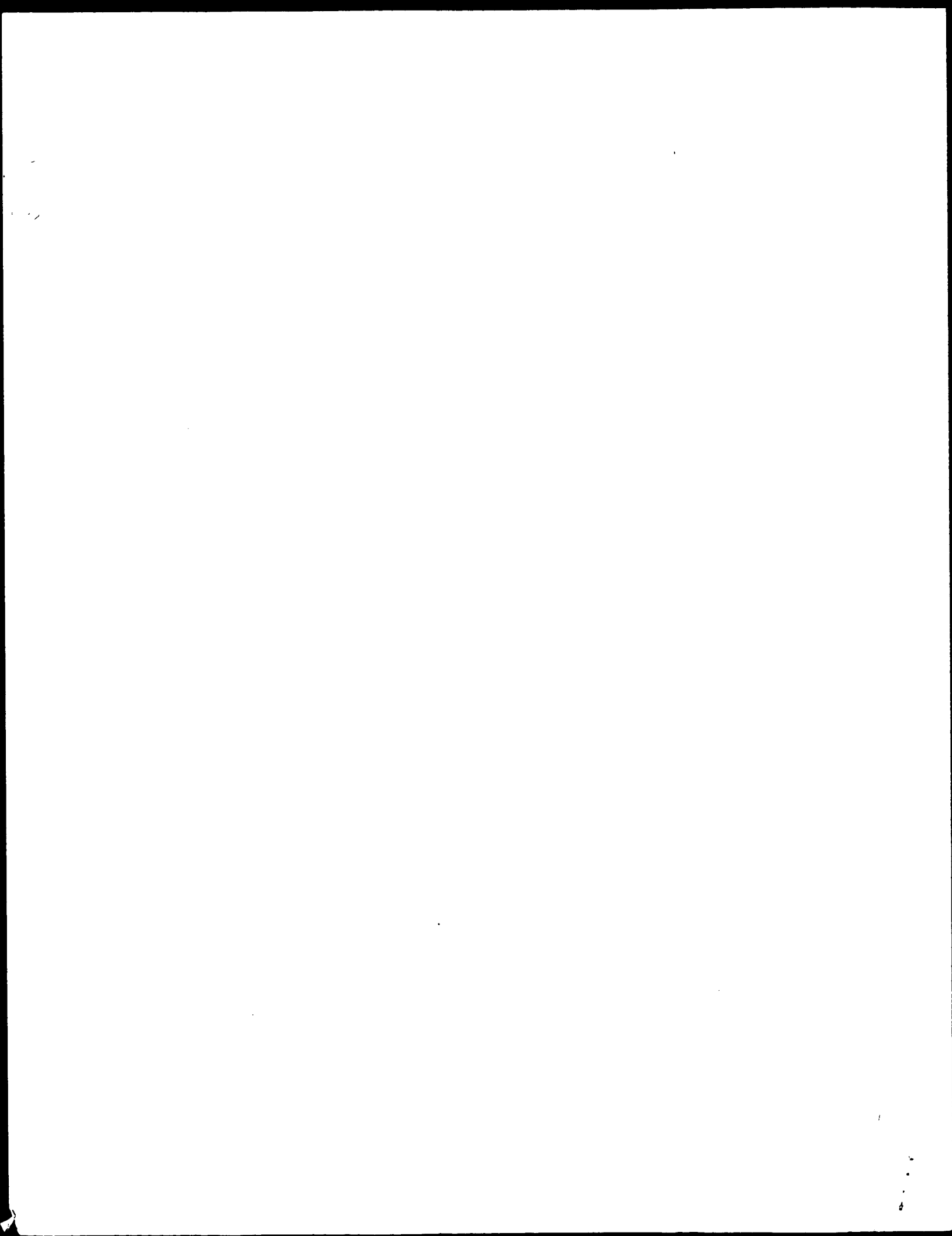
Query Match 58.5%; Score 226.8; DB 19; Length 483;  
Best Local Similarity 75.4%; Pred. No. 1.4e-57;  
Matches 282; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
Qy 15 agatgggaatgacaccgaaatagctcttcgactgctctccgacttcgcaattg 74  
Db 21 AAGAAGATGGTGTGCGGAAATAGCACCGTCGATGCTGTCATCGGACTTCGCAATCTA 80  
Qy 75 gcttcgagctcagcgatgctccacttcgcgccgattggctccacatgacatcatg 134  
Db 81 GCGGCGAGGCGCAATCGGATGATCGATTGGCGCGCAACTGGCTTCACATGGATATTG 140  
Qy 135 gatggcgatttgcctcccaatttaactattgctcctccagttatgaaagtttgaaaag 194  
Db 141 GACGGCATTTTGTCCCTAATCTTAGGATTGGTGTCTCTGTCATCGAGATTGCGAAG 200  
Qy 195 cacacaaaggcatatttggattgtcaccttatggttaacaaatcctcttgattatgttaa 254  
Db 201 CACACTAATAGCATATCTTGTATGCCACCTATGTTGACGAACCCCATGGATTACGTTGCA 260  
Qy 255 ccttggcgaagcgtggtcttctggttttaacatttcacgttagacatcaaaagataac 314  
Db 261 CAGATGGCTTAAAGCTGGGGCTTCTGTGTTTCACATTCCAGCTCGAGTGGCCCCAAGATA 320  
Qy 315 tgggaagaacttatccaaagaatcaagtcacatggcattgattcctggttagcattaaag 374  
Db 321 TGGCAACAACTTGTGAGAGAGATTAACTACTTGGGATGAGACAGGCGTGGCTCTTAAG 380  
Qy 375 cctgggaccccgct 388  
Db 381 CCTGGAACACCTGT 394

RESULT 14  
AV554048  
LOCUS  
DEFINITION EST248362 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
CLEDI9K11, mRNA sequence.  
ACCESSION AV554048  
VERSION AV554048.1 GI:4385394

KEYWORDS	EST.	AI772355	AI772355	552 bp	mrna	EST	29-JUN-1999
SOURCE	tomato.	LOCUS	EST253455	tomato resistant,	Cornell Lycopersicon esculentum	cdna	
ORGANISM	Lycopersicon esculentum	DEFINITION	clone CLER2M16, mRNA sequence.				
REFERENCE	1 (bases 1 to 497)	ACCESSION	AI772355				
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D., and Giovannoni,J.	VERSION	AI772355.1	GI:5270396			
TITLE	Generation of ESTs from tomato carpel tissue	KEYWORDS	EST.				
JOURNAL	Unpublished (1999)	SOURCE	tomato.				
COMMENT	Contact: David Frisch Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU	ORGANISM	Lycopersicon esculentum				
FEATURES	Location/Qualifiers	REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.				
source	1. .497	AUTHORS	1 (bases 1 to 552) D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D., Giovannoni,J.J. and Martin,G.B.				
	/organism="Lycopersicon esculentum"	TITLE	Generation of ESTs from Pseudomonas resistant tomato				
	/cultivar="TA496"	JOURNAL	Unpublished (1999)				
	/db_xref="taxon:4081"	COMMENT	Contact: David Frisch Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU				
	/clone="cLED19K11"	FEATURES	5 prime sequence.				
	/clone_lib="tomato ovary, TAMU"	source	Location/Qualifiers				
	/tissue_type="carpel"		1. .552				
	/dev_stage="5 days pre-anthesis to 5 days post-anthesis"		/organism="Lycopersicon esculentum"				
	/lab_host="X11-Blue MRF"		/cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"				
	/notes="Vector: pBlueScript SK(-); Site.1: EcoRI; Site.2: XhoI; cLED - Tomato Carpel EST Library. OligoRT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."		/db_xref="taxon:4081"				
			/clone="CLER2M16"				
			/clone_lib="tomato resistant, Cornell"				
			/tissue_type="leaf"				
			/dev_stage="4-week old"				
			/lab_host="SOLR"				
			/note="Vector: pBlueScript SK(-); Site.1: EcoRI; Site.2: XhoI; cLER - Tomato Pseudomonas Resistant EST Library. Directionally cloned cDNAs inserted into pBlueScript SK(-) at 5' end with EcoRI and 3' end with XhoI site."				
BASE COUNT	138 a 114 c 125 g 120 t						
ORIGIN							
Query Match	58.2%; Score 226; DB 11; Length 497;						
Best Local Similarity	74.9%; Pred. No. 2.4e-57;						
Matches	283; Conservative 0; Mismatches 95; Indels 0; Gaps 0;						
QY	8 gaaagaaagatgggaatgacacccgaaatagctcttcgactgctcttcgacttcgc 67						
Db	78 GCAAGTGAGGAGATGGTGAAGGCAATCATAGCACCGTCGATGCTCATCGGACTTGC 137						
QY	68 caatttgcttcgaggtcagcgcatgctccacttcggcgccgattggctccacatgga 127						
Db	138 TAATTTGGCTCTGAAGCAGAGCGCATGCTCAATTGTGTCGGATTGGCTCCATATGGA 197						
QY	128 catcatgagtggtggaattgtcccaatttaactattggcgctcagttattgaaattt 187						
Db	198 CATCATGGAGGTGCTATTTGTCCTCCAAACCTTACCGTCGGTCTCCAAATATCGAGAGTCT 257						
QY	188 gagaagacacaaaggcatatttgattgtcaaccttatggttacaaatcctcttgatta 247						
Db	258 GAGAAGCATACAAAGGATATCTGGACTGCCACCTATATGTCACCAACCCCTTGATTA 317						
QY	248 tgttgaaccttggcaaaagctgggtgtctgtgtttacattacacgtagagacatacaa 307						
Db	318 TGTGGAACCGCTTGGCAAAGCTGGTGCTCAGGCTTTACGTTCCATATTAGGACATCTAG 377						
QY	308 agataactggaagaaccttaccagaatcaagtcacatgcatgattcctggtgtagc 367						
Db	378 AGATAATTGGCAAGCTTGTTCACAGGATAAAGTCTAAGGCATGAACCTGGGGTTTC 437						
QY	368 attaaagctgggacccc 385						
Db	438 TTTGAAGCCTGTACACC 455						
RESULT	15						

QY 368 attaaagcctgggacccc 385  
|| ||||| || ||  
Db 406 TTGAAGCCTGGTACACC 423

Search completed: November 4, 2000, 11:52:13  
Job time: 18250 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:44:57 ; Search time 320.8 Seconds  
(without alignments)  
370.042 Million cell updates/sec

Title: US-09-300-482-225

Perfect score: 316

Sequence: 1 gataaggtgcacactgag.....gaatggaggagtagcttoga 316

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	46.2	14.6	1246	21 229148	Corn zeaxanthin ep
2	44.2	14.0	1050	21 245413	D-ribulose-5-phosp
3	35.8	11.3	1260	21 245412	D-ribulose-5-phosp
4	34.4	10.9	533	16 T20487	Human gene signatu
C 5	34.4	10.9	613	21 280305	Human colon cancer
C 6	34.4	10.9	667	21 280304	Human colon cancer
C 7	34.4	10.8	2930	17 T42174	Maize id gene. Ze
C 8	34.4	10.8	3669	19 V41721	Maize id gene. Ze
C 9	34.4	10.8	3693	20 Z34479	Maize id gene. Ze
C 10	33.8	10.7	2639	17 T06481	Cystathionine gamm
C 11	33.4	10.6	5511	20 Z08718	Chimeric gene cons
12	32.8	10.4	1149	20 Z21189	Zea mays pathogene

13	32	10.1	500	19 V63934	Mycobacterium tube
14	32	10.1	500	20 X81041	Nucleotide sequenc
C 15	32	10.1	2050	19 V63935	Mycobacterium tube
C 16	32	10.1	2050	20 X81042	Nucleotide sequenc
C 17	32	10.1	16885	17 T33535	BCG deletion regio
18	31.8	10.1	5511	20 Z08718	Polynucleotide seq
C 19	31.4	9.9	9848	20 X20503	Chimeric gene cons
C 20	31	9.8	341	21 A00751	Human 5' EST isola
C 21	30.2	9.6	300	21 Z57864	Human colon cancer
C 22	30.2	9.6	6138	21 X57864	Protein regulating
C 23	29.8	9.4	622	18 Z30872	Streptococcus pneu
C 24	28.2	9.2	1149	20 Z21189	Zea mays pathogene
C 25	28.8	9.1	8585	11 Q05749	Microspore-specifi
C 26	28.8	9.1	8585	18 T48867	Brassica napus mic
C 27	28.8	9.1	17082	20 X13166	Enterococcus faeca
C 28	28.4	9.0	116277	20 X20249	Borrelia burgdorfe
C 29	28.2	8.9	3475	13 Q21453	Sequence encoding
C 30	28.2	8.9	10329	21 Z50264	Genomic DNA of pot
C 31	28.2	8.9	10329	21 Z50264	Genomic DNA encodi
C 32	28.2	8.9	12566	19 V52282	Streptococcus pneu
C 33	28	8.9	8793	21 Z89350	Murine embryo macr
C 34	27.8	8.8	232	20 V90238	EST clone DF971.
C 35	27.8	8.8	249	8 N70901	ORF 8 from MYMV co
C 36	27.8	8.8	2723	8 N70893	MYMV strand (b).
C 37	27.8	8.8	3305	21 A26925	Essential Staphylo
C 38	27.8	8.8	4360	18 T44068	Rat von Ebner's gl
C 39	27.8	8.8	5836	16 Q96296	Human prostate-spe
C 40	27.8	8.8	5836	18 T84444	Human prostate-spe
C 41	27.8	8.8	5836	21 Z95932	DNA sequence of co
C 42	27.8	8.8	7646	18 V74432	Staphylococcus aur
C 43	27.6	8.7	997	13 Q29266	Human calcium chan
C 44	27.6	8.7	1803	13 Q26409	Toxoplasma gondii
C 45	27.4	8.7	756	20 X98830	Human validated ca

#### ALIGNMENTS

##### RESULT 1

229148/c  
ID 229148 standard; cDNA; 1246 BP.

XX AC 229148;

XX 21-FEB-2000 (first entry)

XX Corn zeaxanthin epoxidase C-terminal half encoding cDNA.

XX DE Corn; zeaxanthin epoxidase; carotenoid biosynthetic enzyme;  
XX KW chimeric gene; transformed cell; inhibitor; antibody; ds.

XX OS Zea mays.

XX FH Key Location/Qualifiers

XX CDS 2..949

XX FT /\*tag= a

XX FT /product= "zeaxanthin epoxidase C-terminal half"

XX PN WO9955889-A2.

XX PD 04-NOV-1999.

XX PF 22-APR-1999; 99WO-US08789.

XX PR 24-APR-1998; 98US-0083042.

XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.

XX PI Cahoon RE, Hitz WD, Shen JB, Williams ME;

XX DR WPI: 2000-072184/06.

XX DR P-PSDB; Y44220.



1

PT Novel nucleic acids, used to develop products for the diagnosis and  
PT treatment of disorders involving unwanted cell proliferation,  
PT particularly cancers, especially colon cancer  
XX  
XX  
XX Claim 15; Page 287; 469pp; English.  
XX  
XX Z79917 to 280766 represent double stranded cDNA clones isolated from the  
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
CC cDNA clones can be used to generate antisense oligonucleotides which  
CC can be used for antisense therapy. Methods and products from the present  
CC invention can be used for identifying and/or classifying cancerous cells  
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas  
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used  
CC for developing agents for the diagnosis and treatment of disorders  
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or  
CC hyperplasia.  
XX  
SQ Sequence 613 BP; 188 A; 151 C; 128 G; 137 T; 9 other;  
  
Query Match 10.9%; Score 34.4; DB 21; Length 613;  
Best Local Similarity 63.1%; Pred. No. 0.069;  
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
  
QY 57 tggctaggtcctcaaccatagacgtgcccgcacatctgctggggcccaattgcctgc 116  
Db 354 TGGTGCAGGTGCTGCAACAGTAGGAGTGGTCTGCTGGTATTGGAAACAGTCTT 295  
  
QY 117 tggagctcttatattgctgc 140  
Db 294 TGGCAGCCTTATCATGTTATGC 271  
  
RESULT 6  
Z80304/c  
ID Z80304 standard; cDNA; 667 BP.  
XX  
AC Z80304;  
XX  
DT 07-APR-2000 (first entry)  
XX  
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:388.  
XX  
XX Human; gene expression product; diagnosis; tumour; colon cancer;  
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
KW hyperplasia; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO9964576-A2.  
XX  
PD 16-DEC-1999.  
XX  
PF 09-JUN-1999; 99WO-IB01062.  
XX  
PR 10-JUN-1998; 98US-0088801.  
XX  
PA (FARB ) BAYER CORP.  
XX  
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;  
PI Schlegel R;  
XX  
DR WPI; 2000-087220/07.  
XX  
XX Novel nucleic acids, used to develop products for the diagnosis and  
PT treatment of disorders involving unwanted cell proliferation,  
PT particularly cancers, especially colon cancer  
XX  
XX Claim 15; Page 286-287; 469pp; English.  
XX  
CC Z79917 to 280766 represent double stranded cDNA clones isolated from the

CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
CC cDNA clones can be used to generate antisense oligonucleotides which  
CC can be used for antisense therapy. Methods and products from the present  
CC invention can be used for identifying and/or classifying cancerous cells  
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas  
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used  
CC for developing agents for the diagnosis and treatment of disorders  
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or  
CC hyperplasia.  
XX  
SQ Sequence 667 BP; 192 A; 171 C; 153 G; 150 T; 1 other;  
  
Query Match 10.9%; Score 34.4; DB 21; Length 667;  
Best Local Similarity 63.1%; Pred. No. 0.071;  
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
  
QY 57 tggctaggtcctcaaccatagacgtgcccgcacatctgctggggcccaattgcctgc 116  
Db 352 TGGTGCAGGTGCTGCAACAGTAGGAGTGGTCTGCTGGTATTGGAAACAGTCTT 293  
  
QY 117 tggagctcttatattgctgc 140  
Db 292 TGGCAGCCTTATCATGTTATGC 269  
  
RESULT 7  
T42174/c  
ID T42174 standard; DNA; 2930 BP.  
XX  
AC T42174;  
XX  
DT 06-MAR-1997 (first entry)  
XX  
DE Maize Id gene.  
XX  
KW Maize; Zea mays; Id; id\*; transposon; transposable element;  
KW Ds2; flower evocation; zinc-finger regulatory protein; ss.  
XX  
XX Zea mays.  
XX  
FH Key Location/Qualifiers  
FT CDS (complement)1..410  
FT /\*tag= a  
FT /note= "the DNA strand given in T42175 is  
FT complementary to this fragment"  
FT intron 176..1600  
FT /\*tag= b  
FT prim\_transcript 1..1890  
FT /\*tag= c  
FT /note= "nucleotides 1 through at least 1890  
FT are transcribed (possibly through 2150)"  
FT prim\_transcript 1..2150  
FT /\*tag= d  
FT /note= "nucleotides 1 through at least 1890  
FT are transcribed (possibly through 2150)"  
FT misc\_RNA 168  
FT /\*tag= e  
FT /note= "the Ds2 transposon insertion occurs  
FT at nucleotide 168"  
XX  
XX WO9634088-A2.  
XX  
XX 31-OCT-1996.  
XX  
XX 15-MAR-1996; 96WO-US03466.  
XX  
XX 16-MAR-1995; 95US-0406186.  
XX  
XX (COLD-) COLD SPRING HARBOR LAB.  
XX  
XX Colasanti JJ, Sundaresan V;  
XX



```

DR WPI; 1996-497621/49.
XX P-PSDB; W03698.
PT New isolated plant Id gene - used to develop prods. for use in
PT altering the induction of flowering in plants
XX Claim 6; Fig 2 and Fig 3; 58pp; English.
XX
CC The Id gene controls flower evocation in maize plants. The
CC maize nucleic acid is similar to that of genes encoding zinc-
CC finger regulatory proteins in animals.
CC Transposons Ac and Ds constitute a family of related transposable
CC elements present in maize. A derivative of Ds, Ds2, can be used
CC to produce a new mutant of the Id gene. The Ds2 (in the presence
CC of active Ac) is excised from a nearby gene on chromosome 1
CC and inserted into the Id gene to produce Id*.
XX
XX Sequence 2930 BP; 850 A; 664 C; 643 G; 771 T; 2 other;
SQ
Query Match 10.8%; Score 34; DB 17; Length 2930;
Best Local Similarity 74.1%; Pred. No. 0.19;
Matches 43; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 235 ccggttttttttattcgctgcttttagttcaaacatgaactagcgacgactgat 292
Db 710 CCTTCTTTTTTATCAGTGGCGCATTTAGTTTAAAAATCACTAGCGGACGATAAATAT 653
RESULT 8
ID V41721/c
XX V41721 standard; DNA; 3669 BP.
XX AC V41721;
XX
XX 16-NOV-1998 (first entry)
DE Maize Id gene.
XX
XX ss; maize; Id; floral induction; transgenic plant.
XX
XX Zea mays.
FH Key Location/Qualifiers
FT CDS 12..2961
FT exon /*tag= a
FT /*product= "Id protein"
FT 12..228
FT /*tag= b
FT /number= 1
FT 229..318
FT /*tag= c
FT /number= 1
FT exon 319..609
FT /*tag= d
FT /number= 2
FT misc_feature 380..442
FT /*tag= e
FT /note= "Zinc finger motif"
FT intron 610..728
FT /*tag= f
FT exon 729..902
FT /*tag= g
FT /number= 3
FT misc_feature 796..858
FT /*tag= h
FT /note= "Zinc finger motif"
FT intron 903..2327
FT /*tag= i
FT /number= 3
FT exon 2328..2959
FT /*tag= j

```

---

```

FT /number= 4
FT 3179
FT /*tag= k
PN WO9837201-A1.
XX PD 27-AUG-1998.
XX
XX 18-FEB-1998; 98WO-US03161.
XX
XX 30-DEC-1997; 97US-0000640.
XX 20-FEB-1997; 97US-0804104.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX
XX Colasanti JJ, Sundaresan V;
XX
XX WPI; 1998-467564/40.
XX P-PSDB; W59836.
XX
XX Id gene controlling floral induction in maize - useful to create
XX transgenic plants with earlier, delayed or inhibited floral
XX induction e.g. to extend geographical range of crops
XX
XX Claim 1; Fig 2; 68pp; English.
XX
XX The maize Id gene controls floral induction and is useful in the
XX production of transgenic plants with altered floral induction times,
XX i.e. earlier, delayed or inhibited floral induction. The Id gene and
XX zinc-finger regulatory regions are useful to identify these genes in
XX maize and isolate similar genes in other plants. The polypeptides
XX and antibodies are also useful in Id detection e.g. to locate activity in
XX plants.
XX
XX Sequence 3669 BP; 1011 A; 922 C; 851 G; 884 T; 1 other;
SQ
Query Match 10.8%; Score 34; DB 19; Length 3669;
Best Local Similarity 74.1%; Pred. No. 0.21;
Matches 43; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 235 ccggttttttttattcgctgcttttagttcaaacatgaactagcgacgactgat 292
Db 1437 CCTTCTTTTTTATCAGTGGCGCATTTAGTTTAAAAATCACTAGCGGACGATAAATAT 1380
RESULT 9
ID Z34479/c
XX Z34479 standard; DNA; 3693 BP.
XX AC Z34479;
XX
XX 01-FEB-2000 (first entry)
XX
XX Maize Id gene.
XX
XX Indeterminate gene; Id gene; maize; corn; flower induction;
XX floral evocation; transgenic plant; ss.
XX
XX Zea mays.
FH Key Location/Qualifiers
FT CDS 12..2957
FT /*tag= a
FT /note= "contains introns"
FT exon 12..240
FT /*tag= b
FT /number= 1
FT intron 241..329
FT /*tag= c
FT /number= 1
FT exon 330..627
FT /*tag= d

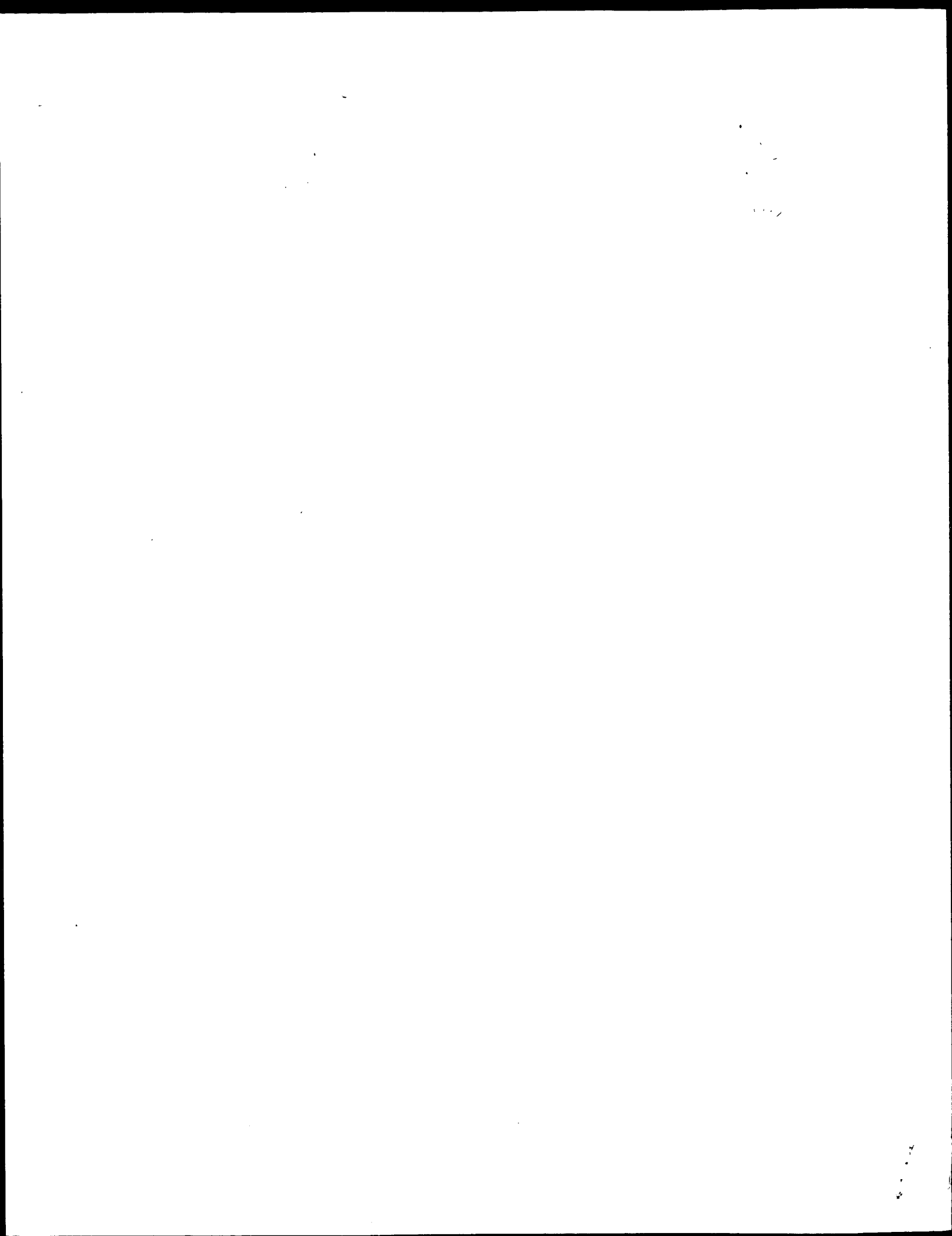
```







```
XX Key Location/Qualifiers
FH CDS 22..2022
FT /tag- a
FT
XX WO9844119-Al.
XX
XX PD 08-OCT-1998.
XX
XX PF 01-APR-1998; 98WO-DK00132.
XX
XX PR 05-JAN-1998; 98US-0070488.
XX
XX PR 02-APR-1997; 97DK-0000376.
XX
XX PR 18-APR-1997; 97US-0044624.
XX
XX PR 10-NOV-1997; 97DK-0001277.
XX
XX (STAT-) STATENS SERUM INST.
XX
XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;
PI Rosenkrands I, Weldingh K;
XX
XX WPI: 1998-542705/46.
XX P-PSDB; W72911.
XX
XX New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis
XX
XX Claim 23; Page 180-183; 163pp; English.
XX
XX The present sequence encodes a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
XX
XX Sequence 2050 BP; 404 A; 774 C; 582 G; 290 T; 0 other;
SQ
Query Match 10.1%; Score 32; DB 19; Length 2050;
Best Local Similarity 55.4%; Pred. No. 0.74;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 31 ccttccttgacatagaggttggtgtcttagcttcacacatagacgtgcccga 90
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 CCGCGCATGGCATGGGGGTGGGTGGTTAGATCGGCCCGTTCGCGAGGGCGGC 279
QY 91 tctgctggggccaattgcatcgtcgtggaagctctatatattgctgctg 142
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 TCTCTGCGCGCATGGCATGGAGTTGGCGAGGGAGGTGGGGTGGGG 227
Search completed: November 4, 2000, 13:45:01
Job time: 16428 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:33:34 ; Search time 189.35 Seconds  
(without alignments)  
252.402 Million cell updates/sec

Title: US-09-300-482-225  
Perfect score: 316  
Sequence: 1 gataaggtcgacactgag.....gaatggaggagtagtacttoga 316

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.8	10.7	3639	4	US-08-737-524B-26
2	32	10.1	16885	2	US-08-390-878-16
3	29.2	9.2	6343	5	US-08-581-148C-30
4	28.8	9.1	8595	1	US-08-030-096-3
5	28.2	8.9	3475	2	US-07-960-389-1
6	27.8	8.8	3305	5	US-08-714-918-76
7	27.8	8.8	4360	1	US-08-470-350B-1
8	27.8	8.8	5836	1	US-08-380-916-1
9	27.8	8.8	5836	5	US-08-721-690-1
10	27.6	8.7	1803	5	US-08-458-922-2
11	27.2	8.6	2375	1	US-08-468-853-9
12	27.2	8.6	2375	2	US-08-468-855-9
13	27.2	8.6	2375	2	US-08-310-357-9
14	27.2	8.6	2375	2	US-08-468-852-9
15	27.2	8.6	2375	4	US-08-468-857-9
16	27	8.5	5035	4	US-08-616-392C-3
17	26.8	8.5	1950	4	US-08-377-440A-2
18	26.8	8.5	3765	5	US-07-705-490-1
19	26.8	8.5	4362	3	US-08-455-073A-1
20	26.8	8.5	28804	3	US-08-592-874-1
21	26.8	8.5	28804	5	US-09-096-942-2
22	26.8	8.5	28804	5	US-09-096-867-2
23	26.8	8.5	29604	5	US-08-781-891-207
24	26.6	8.4	572	2	US-08-253-155A-15
25	26.4	8.4	2631	2	US-08-785-429-1
26	26.4	8.4	2631	5	US-08-996-621-1

c 27	26.4	8.4	40352	5	US-08-846-111D-15	Sequence 15, Appl
c 28	26	8.2	2688	1	US-08-088-633-3	Sequence 3, Appl
c 29	26	8.2	2688	1	US-08-245-756-3	Sequence 3, Appl
c 30	26	8.2	2688	2	US-08-441-750-3	Sequence 3, Appl
c 31	26	8.2	2688	3	US-08-441-751-3	Sequence 3, Appl
c 32	26	8.2	2688	6	PCT-US92-02521-3	Sequence 3, Appl
c 33	25.8	8.2	1478	5	US-08-817-926-1	Sequence 1, Appl
c 34	25.8	8.2	2322	2	US-08-618-164-1	Sequence 1, Appl
c 35	25.8	8.2	3562	5	US-08-817-926-19	Sequence 1, Appl
c 36	25.6	8.1	428	5	US-08-755-587-10	Sequence 10, Appl
c 37	25.6	8.1	501	5	US-08-755-587-9	Sequence 9, Appl
c 38	25.6	8.1	1095	3	US-08-180-524-2	Sequence 2, Appl
c 39	25.6	8.1	1095	4	US-08-975-166-2	Sequence 2, Appl
c 40	25.6	8.1	1129	5	US-08-755-587-30	Sequence 30, Appl
c 41	25.6	8.1	1540	2	US-08-463-115-2	Sequence 2, Appl
c 42	25.6	8.1	1540	2	US-08-465-388-2	Sequence 2, Appl
c 43	25.6	8.1	2381	1	US-08-021-608D-9	Sequence 9, Appl
c 44	25.6	8.1	2381	2	US-08-726-160-9	Sequence 9, Appl
c 45	25.6	8.1	2381	6	PCT-US94-01782-9	Sequence 9, Appl

## ALIGNMENTS

RESULT 1  
US-08-737-524B-26  
; Sequence 26, Application US/08737524B  
; Patent No. 5912414  
; GENERAL INFORMATION:  
; APPLICANT: CARL SAVERIO FALCO  
; APPLICANT: DOMINICK ANTHONY GUIDA, JR.  
; APPLICANT: MARY ELIZABETH HARNETT LOCKE  
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS, CHIMERIC  
; TITLE OF INVENTION: GENES AND METHODS FOR INCREASING  
; TITLE OF INVENTION: THE METHIONINE CONTENT OF THE SEEDS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/737,524B  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LYNN M. CHRISTENBURY  
; REGISTRATION NUMBER: 30,971  
; REFERENCE/DOCKET NUMBER: BB-1059-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-5481  
; TELEFAX: 302-773-0164  
; TELEX: 835420  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3639 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-737-524B-26

Query Match 10.7%; Score 33.8; DB 4; Length 3639;  
Best Local Similarity 68.1%; Pred. No. 0.033;





APPLICATION NUMBER: PCT/US 91/04846



	Query Match	8.8%;	Score 27.8;	DB 5;	Length 5836;
	Best Local Similarity	53.2%;	Pred. No. 6.1;		
	Matches	59;	Conservative	0;	Mismatches 52; Indels 0
QY	166	ctgagagaagcgcgcgaggcctctcagacacaaactgatttggctgtttctgcgcg			
Db	4839	CTGGGTACAGAGTGAGACTCTCTCAAAAAAATTTTTTTTTTTTTTTTGG			
QY	226	gtactccctccgtttttttttattgcgcgcgttttagttcaaacatgaact			
Db	4899	GATGGATCTGTGCTTTGTTCTCTGGTTGGCCVTGAACCTCTGGCTTCAAGT			

[illegible]

```

RESULT 11
US-08-468-853-9
; Sequence 9, Application US/08468853
; Patent No. 5670362
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5670362el Patent Department
;

```

```

RESULT 11
US-08-468-853-9
; Sequence 9, Application US/08468853
; Patent No. 5670362
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAERT, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5670362el Patent Department
;

```

STREET: 1300 Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,853  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/310,357  
FILING DATE: 21-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/102,865  
FILING DATE: 06-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/604,075  
FILING DATE: 18-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91.201.523.7  
FILING DATE: 18-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary E. Gormley  
REGISTRATION NUMBER: 34,409  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2375 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
IMMEDIATE SOURCE:  
CLONE: Eam100E  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2-1950

		Query Match	8.6%;	Score 27.2;	DB 1;	Length 2375;
		Best Local Similarity	52.7%;	Pred. No. 6.1;		
		Matches	59;	Conservative	0;	Mismatches 53; Indels 0; Gaps 0;
QY	47	aggttgatggtggtctctaggctctcaacatagacgctgcgcgcctctgctggggccaatt	106			
Db	1576	AGATTGAAGCTGTTGCAGGACGTCACACGAAAAAACAAGCGCAAAATGTCTGGATCTAAGG	1635			
QY	107	gcatgcgcctcggaagctctctatatattggcgcctgcggaccaccgagagcatcat	158			
Db	1636	TAGGCTGCGCTCGTAGTCTAGAGGTGGCTATGCTTTATAAGGAAGACCTCCTT	1687			

```

RESULT 12
US-08-468-855-9
; Sequence 9, Application US/08468855
; Patent No. 5780289
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5780289el Patent Department
; STREET: 1300 Piccard Drive
;

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:40:30 ; Search time 320.8 Seconds  
(without alignments)  
480.118 Million cell updates/sec

Title: US-09-300-482-14

Perfect score: 410

Sequence: 1 cccacgcgtccgcgggtcatg.....gcaacggcccgctcttgatg 410

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_36:\*

- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	99.2	24.2	1916	18	T88035
2	91.2	22.2	1038602	20	Z01425
3	69.2	16.9	14516	20	X06748
4	69.2	16.9	14516	21	Z56375
5	66.2	16.1	4487	7	N60943
6	65.6	16.0	2450	20	X20537
7	57	13.9	2025	19	V42965
8	57	13.9	2025	20	X16683
9	56.4	13.8	1953	19	V52287
10	56.4	13.8	2087	20	Z30047
11	49.6	12.1	16592	18	V74364
12	39.2	9.6	53789	19	V21187

13	37.4	9.1	1524	18	T63571	Xylanase gene obta
14	35.6	8.7	939	20	X33560	Rice beta-glucanas
15	35.6	8.7	1020	20	X33570	Rice Gns8 CDS. Or
16	35.6	8.7	2088	12	Q14479	Truncated HSVgB ge
17	35.6	8.7	2612	20	X33536	Rice beta-glucanas
18	35.6	8.7	2925	12	Q14455	HSV surface antige
19	35.6	8.7	3465	12	Q14478	HSVgB gene. Herpe
20	35.6	8.7	8051	18	T2685	Sugar biosynthesis
21	35.6	8.7	8160	20	X25774	S.erythraea erythr
22	35.4	8.6	811	21	Z99478	Maize gibberellin
23	35.2	8.6	2298	20	X06781	Human testis secre
24	35.2	8.6	2990	19	V66832	zea mays soluble s
25	35.2	8.6	3476	18	T95698	Novel human gene,
26	34.8	8.5	1053	10	N90184	DNA encoding to 6-
27	34.8	8.5	1411	13	Q30672	Homo sapiens mamma
28	34.4	8.4	2277	19	V13836	Human telomerase p
29	34.4	8.4	2277	19	V03372	F. lutescens piper
30	34.2	8.3	6357	21	Z91051	Herpes Simplex vir
31	34	8.3	2712	8	N71303	Glycoprotein B (gB
32	34	8.3	3472	14	Q48497	Herpes simplex vir
33	34	8.3	3473	18	T93651	Sequence of Herpes
34	34	8.3	3474	9	N80908	Sequence encoding
35	34	8.3	3966	6	N50516	Human retinal dege
36	34	8.3	5019	19	V34855	Streptomyces livid
37	33.8	8.2	1910	14	Q34549	Leishmania brazili
38	33.2	8.1	2040	18	T62137	Leishmania antigen
39	33.2	8.1	2040	19	V47557	Gene encoding a su
40	33.2	8.1	6741	21	A10595	A. crysozenum cyst
41	33	8.0	1778	16	Q86722	Sequence of herpes
42	33	8.0	2426	6	N50490	Herpes Simplex vir
43	33	8.0	2426	10	N90525	A. crysozenum cyst
44	33	8.0	3494	16	Q89311	Mycobacteriophage
45	33	8.0	49272	19	V35000	

#### ALIGNMENTS

RESULT 1	
T88035	
ID	T88035 standard; DNA; 1916 BP.
XX	
AC	T88035;
XX	
DT	22-DEC-1997 (first entry)
XX	
DE	Brevibacterium flavum 6-phosphogluconate dehydrogenase DNA.
XX	
KW	Brevibacterium flavum; 6-phosphogluconate dehydrogenase;
KW	recombinant production; coryneform; bacterium; ds.
XX	
OS	Brevibacterium flavum.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	374..1852
FT	/*tag= a
FT	/product= 6-phosphogluconate_dehydrogenase
XX	
PN	JP09224662-A.
XX	
PD	02-SEP-1997.
XX	
PF	23-FEB-1996; 96JP-0036346.
XX	
PR	23-FEB-1996; 96JP-0036346.
XX	
PA	(MITU ) MITSUBISHI CHEM CORP.
XX	
DR	WPI: 1997-484097/45.
DR	P-PSDB; W27613.
XX	
PT	6-phospho:gluconate dehydrogenase - can be recombinantly produced by transforming coryneform bacteria with DNA molecule encoding it
PT	

```

XX PS Claim 1; Pages 5-7; 8pp; Japanese.
XX CC The present sequence encodes the Brevibacterium flavum JM-233
XX CC 6-phosphogluconate dehydrogenase (6-PD). The 6-PD can be
XX CC recombinantly produced by transforming coryneform bacteria with the
XX CC DNA molecule encoding it.
XX SQ Sequence 1916 BP; 425 A; 596 C; 507 G; 388 T; 0 other;

Query Match          24.2%; Score 99.2; DB 18; Length 1916;
Best Local Similarity 54.9%; Pred. No. 5.3e-17;
Matches 219; Conservative 0; Mismatches 174; Indels 6; Gaps 1;

QY 12 qcgggtcatggggcagaaccttgcctcaacattgcagagaaagggttccccatctctgtg 71
Db 443 gcagtaatgggtcaaaccttcgcccgaacttcgcgcgaacggccacactgtcgtctc 502

QY 72 tacaacaggacaaacctcaacaggtggacgagaccgtgcagcgtgcccgaaggcagaagaaac 131
Db 503 tacaacgcagcactgacaaacgcgaacgaagctcatcgcgcgcatcagcggctccgaaggcaac 562

QY 132 ctcccgctcagcgtctccatgaccccgctcctttgtgaagtccattcagaagccaacgg 191
Db 563 ttcatccttcgcaacgcgtcgaagag-----ttcgtagcatcctcgtgaaagccaacgc 616

QY 192 gtgggtgatcatcgtcgtcaagcgccgcgcgcaggttaccagacaccatcggaogctcgca 251
Db 617 cgcgcacatcatcattggttcaggctggttaacgcaccgcagcagtcacacacagctggca 676

QY 252 gtcacttggagcagggcagactgatacatcgtatcgatggggggaacagagtgtacgaagaacacg 311
Db 677 gacgcacatggacgaaggcgacatcatcgtacgcagcgcgcgaacgcctctctacacgcgaccc 736

QY 312 gagagaggaggaagccatgagagagcgcgcctnctgtatcttggcattgggtgtctct 371
Db 737 attctcgcgcgagaggaataatccgcagcgcggtctccacttcgtcgtggtgtatctct 796

QY 372 ggaggaaggagggtggccgcgaacggccgcctcttgatg 410
Db 797 ggcgcggaagaaggcgactcaacggccacatccatcatg 835

RESULT 2
Z01425/c
ID Z01425 standard; DNA; 1038602 BP.
XX AC Z01425;
XX DT 07-OCT-1999 (first entry)
XX DE Complete genome sequence of Chlamydia trachomatis.
XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
XX KW nonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX KW bartholinitis; pneumonia; venereal lymphogranulomatosis; ss.
XX OS Chlamydia trachomatis.
XX PN WO9928475-A2.
XX PD 10-JUN-1999.
XX PF 27-NOV-1998; 98WO-TB01939.
XX PR 04-NOV-1998; 98US-0107077.
XX PR 28-NOV-1997; 97FR-0015041.
XX PR 17-DEC-1997; 97FR-0016034.
XX PA (GEST ) GENSET.
XX PA

PI Griffiths R;
XX WPI; 1999-371125/31.
XX PT Genome sequence of Chlamydia trachomatis
XX PS Claim 1; Page 373-656; 1755pp; English.
XX CC The present sequence represents the complete genome of Chlamydia
XX CC trachomatis. Open reading frames (ORFs) of the genome encode polypeptides
XX CC Y3754-Y37949. The polypeptides can be used as vaccines against
XX CC Chlamydia trachomatis. Antisense and ribozyme sequences can also be
XX CC used to control growth of the microorganism. Chlamydia trachomatis is
XX CC responsible for a large number of diseases, e.g. eye diseases such as
XX CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
XX CC conjunctivitis; genital diseases such as nongonococcal urethritis,
XX CC epididymitis, cervicitis, salpingitis, perihhepatitis, bartholinitis;
XX CC pneumonia in breast feeding infants; and venereal lymphogranulomatosis.
XX CC The polypeptides of the invention may be of use in treating these
XX CC diseases.
XX SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match          22.2%; Score 91.2; DB 20; Length 1038602;
Best Local Similarity 53.6%; Pred. No. 3.3e-14;
Matches 214; Conservative 0; Mismatches 179; Indels 6; Gaps 1;

QY 12 qcgggtcatggggcagaaccttgcctcaacattgcagagaaagggttccccatctctgtg 71
Db 397456 GCGGTGATGGGCAAAACCTTGTATTGAACATGGTGGATCATGGTTTTCTGTTTCTGTC 397397

QY 72 tacaacagacaaacctcccaagggtggacgacgcgtgcagcgtgcccgaaggcagaagaaac 131
Db 397396 TATAACCGAAGTCCGCGGAAACACAGAGAGTCTTGAAGATCATGGAGAGTGGAGCT 397337

QY 132 ctcccgctcagcgttccatgaccccgctcctttgtgaagtccattcagaagccaacgg 191
Db 397336 CTGC-----AAGGATTTACTAGATTAAGAGTGTGTTCAATCTTTGAAGCGTCTCGT 397283

QY 192 gtgggtatcatcgtcgtcaagcgccgcgcgcaggttaccagacaccatcgacgctcgca 251
Db 397282 AAGATCATGATCATGATTAAGCGGGAGCTCCGTTGTGTAATGATGCTCCCTGCTT 397223

QY 252 gctcacttgagcagggcgactgcatactcagatggggggaacgagtggtacgagaacacg 311
Db 397222 CCTTCTTGGAGAGGGAGATATTTCTATTGATGGGGGAATAGCTATTATTAGATTCT 397163

QY 312 gagagaggaggagagccatggagagcgcgcctnctgtatcttgcagtggtgtctct 371
Db 397162 GAGCGACGCTATGTCGACCTGAAAAAAGAGGAATTTCTATTGTTGGATGGGAGTCTCT 397103

QY 372 ggaggaaggagggtggccgcgaacggccgcctcttgatg 410
Db 397102 GGAGGGGAAGAGGGGGCTAGAAAAAGGCGCTTCCATTATG 397064

RESULT 3
X06748
ID X06748 standard; DNA; 14516 BP.
XX AC X06748;
XX DT 26-APR-1999 (first entry)
XX DE E. coli O111 antigen gene cluster.
XX KW O antigen; O111 antigen; diagnosis; wbdH gene; wzx gene; wzy gene;
XX KW wbdM gene; rfb gene; transferase; polymerase; flippase; diarrhoea;
XX KW haemorrhagic colitis; ds.
XX OS Escherichia coli.
XX PA

```





PA (UNSY ) UNIV SYDNEY.

XX Reeves PR, Wang L;

XX WPI; 2000-072598/06.

XX Novel nucleic acid molecule useful for the detection of flagellated  
PT bacterial strains in food, faeces, etc. -

XX Claim 3; Page 216-221; 245pp; English.

XX Z56331 to Z56398 represent nucleic acid molecules (I) encoding all or  
CC part of an Escherichia coli flagellin protein except a protein  
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present  
CC invention also describes a method of detecting the presence of E. coli  
CC of a particular H serotype in a sample, comprising specifically  
CC hybridising a nucleic acid, preferably at least a pair, derived from a  
CC flagellating gene, specific for a particular flagellin gene associated  
CC with the H serotype, to any E. coli in the sample which contain the gene,  
CC and detecting any hybridised molecules, identifying the presence of that  
CC serotype in the sample. (I) are useful for: (1) detecting the presence  
CC of E. coli of H serotype in a sample by hybridising at least one or a  
CC pair of (I) to any E. coli in the sample and detecting the hybridised  
CC nucleic acid molecules; and (2) for detecting the presence of both O  
CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)  
CC to any E. coli present in the sample and detecting the hybridised  
CC nucleic acid molecules. (I) is particularly useful for detecting the  
CC combination of O and H antigen. Hybridised (I) when using at least one  
CC (I) is detected by southern blot analysis and, when using a pair of (I),  
CC is detected by polymerase chain reaction (PCR). Z56399 to Z56420  
CC represent primers used in the exemplification of the present invention.

XX SQ Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 other;

Query Match 16.9%; Score 69.2; DB 21; Length 14516;

Best Local Similarity 50.4%; Pred. No. 5.7e-09;

Matches 201; Conservative 0; Mismatches 189; Indels 9; Gaps 1;

QY 12 ggggtcgtgagcagcagccttgcctcaacattgcagagaaagggttcccatctctgtg 71

DB 13157 gcaagtggggcgcaacctggcgtcaacatcgaaagccggttatccgtctccatc 13216

QY 72 tacaacagacaacctccaaggtgacagacccgtgcagcgtgccagggcagaaggaaac 131

DB 13217 tcaacctcctccgcagaaactgaagaagtgtgtccgagaaccgcgataaagaactg 13276

QY 132 cttccgtctacgtgttccatgaccccgctcctcttgtgaagtccattcagaagccacgg 191

DB 13277 gtccctattacacggtgaaga-----gttcgtcgagctcttgaaccaccacgt 13327

QY 192 gtggtgatcatcgtcgtcaagcgcgcgcagtcgacagacacacgcagcgtcgca 251

DB 13328 cgtatcctgttaatgttaaaagcagggcggaactgatgtctatcgtatccctgaag 13387

QY 252 gctcacttgagcagcgcactgcacatcagatgggggaaacagagtggtacgaaacacg 311

DB 13388 ccgtatcttgataaaggagacatcattatgatggtgccaacaccttctccagacact 13447

QY 312 gagagagaggagaagcccatgagagcgcgcgcctnctgtatcttggcatgggtgtctct 371

DB 13448 atccgctgaacccgtgaactgccgcggaaggccttaacttcattcgttacccggcgtgcc 13507

QY 372 gaggaagaggaggtgccgcgaacggcccgctcctgatg 410

DB 13508 ggcggtgaagaggcgccctgtaagggcccatctatcatg 13546

RESULT 5

ID N60943

XX N60943 standard; DNA; 4487 BP.

XX N60943;

XX 07-OCT-1991 (first entry)

XX Sequence encoding gluconic acid promoter and operon.

XX Catabolite repression; ss.

XX Bacillus subtilis.

XX JP61100194-A.

XX 19-MAY-1986.

XX 23-OCT-1984; 84JP-0221192.

XX 23-OCT-1984; 84JP-0221192.

XX (MITK ) MITSUI TOATSU CHEM INC.

XX WPI; 1986-166561/26.

XX Gluconic acid operon and its promoter of Bacillus subtilis - in  
PT DNA fragment, which is controlled by catabolite repression.

XX Claim 2; Page 499; 10pp; Japanese.

XX The sequence may be used in the expression of hetrogenous sequences,  
CC without catabolite repression, inducible with the action of gluconic  
CC acid.

XX Sequence 4487 BP; 1308 A; 991 C; 1065 G; 1123 T; 0 other;

Query Match 16.1%; Score 66.2; DB 7; Length 4487;

Best Local Similarity 50.0%; Pred. No. 2.6e-08;

Matches 198; Conservative 0; Mismatches 189; Indels 9; Gaps 1;

QY 15 gtcatggggcagaaaccttgcctcaacattgcagagaaagggttcccatctctgttac 74

DB 3932 gtaatgggaagcaacatcgcttaaacatggcaataaaggcgaaacgtcgctctcat 3991

QY 75 aacagacaaacctccaaggtgacagacacgtgcagcgtgccagggcagaagaaacatt 134

DB 3992 aa-----ttacaccagagatttaacggaccagcttatccaaaagctgtaggacaa 4042

QY 135 ccggtctacggcttccatgaccccgctcctcttggaaagtcattcagaagccacgggtg 194

DB 4043 tctctcagccgtattacagcgttgagactttgttcaatcgttagaaaaccacagaaa 4102

QY 195 gtgatcatcgtctcaaggccggcgcgcagttgaccagaccatcgacgctcgcagct 254

DB 4103 atctttttaggtgtcacagcgggaaacccgtagattccgtcatccaatcattaaagct 4162

QY 255 caattggagcagggcagctgcatcagatgggggaaacagagtggtacgagaacacggag 314

DB 4163 ttgcttgaagaagcagctcatcattggaaggggaaaccccatcattgaagacagaa 4222

QY 315 aggagggaggaagccatggagagcgcgcctnctgtatcttggcatgggtgtctctgga 374

DB 4223 agaagatatcagcagctgaaggaagaaagggatcgcttaoctggagtcgagcatttccgc 4282

QY 375 ggaagaggggttccgcgaacggcccgctccttgatg 410

DB 4283 ggtgaagtcggtcggttaacagggccttccatcatg 4318

RESULT 6

X20537

ID X20537 standard; DNA; 2450 BP.

XX X20537;

XX 05-MAY-1999 (first entry)



Db 232 cacaccttgacaagggtgatattgattgacggtggaataactttctacaaagatacca 291  
 Qy 313 agaggaggaggagccatgagagagcggcgccctnctgtatcttggcatgggtgtcttg 372  
 Db 292 tccgtcgtaataagaattggcaactcaggtatcaactttatcggtactggagttctg 351  
 Qy 373 gaggaaaggaggtgcccgcaacgcccgtcccttgatg 410  
 Db 352 gtgggaaaaagggtgccccttgagggtcccttctatcatg 389

## RESULT 8

X16683  
 ID X16683 standard; DNA; 2025 BP.

AC X16683;  
 DT 04-MAY-1999 (first entry)

XX Streptococcus pneumoniae response regulator ORF DNA sequence.

XX Streptococcus pneumoniae; response regulator; antibacterial; infection;  
 KW pneumonia; bacteraemia; meningitis; otitis media; conjunctivitis;  
 KW sinusitis; pleural empyema; endocarditis; gene therapy; ss.

OS Streptococcus pneumoniae.

XX Key Location/Qualifiers  
 FH 1403..2023

FT CDS /\*tag= a  
 FT /\*note= "no stop codon given"

XX EP892057-A2.

XX 20-JAN-1999.

XX 10-JUL-1998; 98EP-0305517.

XX 18-JUL-1997; 97US-0053238.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

FA (SMIK ) SMITHKLINE BEECHAM PLC.

XX Biswas S, Kosmatka AL, Shilling LK, Throup J, Wallis NG;  
 PI Zalacain M;

XX WPI; 1999-083574/08.

DR P-PSDB; W94667.

XX New Streptococcus pneumoniae response regulator polypeptides and

PT polynucleotides - useful as diagnostic reagents and for prevention

PT and treatment of Streptococcus pneumoniae infections, especially

PT pneumonia, bacteraemia and meningitis

XX Claim 2; Page 5-6; 43pp; English.

XX The present sequence encodes a bacterial response regulator (RR) protein  
 CC which is a component of the two component signal transduction system  
 CC (TCSTs). RR polynucleotides and proteins are useful for diagnosing  
 CC susceptibility to diseases by detecting mutations or polymorphisms in  
 CC the RR gene or analysing for the presence of amount of RR protein  
 CC expressed in a patient sample. RR PCR probes are useful for diagnosing  
 CC diseases, and can characterise the response of the infectious organism  
 CC to drugs. RR proteins and polynucleotides are also useful for screening  
 CC for antagonists, agonists and drugs against infectious micro-organisms.  
 CC RR agonists and antagonists are bacteriostatic and bacteriocidal  
 CC compounds which can be used in treatment to enhance (agonist) or block  
 CC (antagonist or antisense sequence) RR activity, therefore treating  
 CC microbial diseases, especially Streptococcus pneumoniae diseases  
 CC including otitis media, bacteraemia, conjunctivitis, pneumonia,  
 CC sinusitis, pleural empyema, endocarditis and especially meningitis.

CC Epitopes of RR proteins and polynucleotides are useful immunogens for  
 CC producing anti-RR antibodies for prevention of bacterial infections, and

CC RR polynucleotides can be used in genetic immunisation (gene therapy) to  
 CC prevent infections. RR proteins, polynucleotides and their (ant)agonists  
 CC can prevent adhesion of bacteria to matrix proteins, and are useful for  
 CC use on wounds and body implants to prevent bacterial infection.

XX Sequence 2025 BP; 592 A; 413 C; 463 G; 557 T; 0 other;

Query Match 13.9%; Score 57; DB 20; Length 2025;  
 Best Local Similarity 48.5%; Pred. No. 5.2e-06;  
 Matches 193; Conservative 0; Mismatches 196; Indels 9; Gaps 1;

Qy 13 cggctcatgggcagaaaccttgcctcaacattgcagagaaaggggtcccccattctgtgt 72  
 Db 1 cggtaattgggtcgtaaccttgccttaattgaatcaatcggtgtacacaattgtatct 60

Qy 73 acaacaggacaaacctcccaaggtagcagaccgtgcagcgtgcgaaggcagagaagaaacc 132  
 Db 61 acaaccgtagtaaaagaaaaaacggaagatgtgattgcttgcctcctgaaaagaactttg 120

Qy 133 ttccogtctacggttccatgaaccccgctcttctgtgaagtccattcagaagccacggg 192  
 Db 121 taccagct-----atgacgttgaagtcttgaactcaatgaaaaacctctc 171

Qy 193 tggtagatcatgctgcgaagccgcgcgacgttgaccagaccatcgcgacgtctcgag 252  
 Db 172 gtatcatgtgatggttcaagctgacacctgtacagatgtactatccaagccctcttc 231

Qy 253 ctcaacttgagcagggcactgcacatcgatggggggaacagagtggtacagaaacagg 312  
 Db 232 cacaccttgacaagggtgatattgattgacggtggaaatactttctacaaagatacca 291

Qy 313 agagagaggagaagcccatggagagcgccctnctgtatcttggcatgggtgtctctg 372  
 Db 292 tccgtcgtaataagaattggcaactcaggtatcaactttatcggtactgggttctg 351

Qy 373 gaggaaaggaggtgcccgcaacgcccgtccttgatg 410

Db 352 gtgggaaaaagggtgcccttgaaggtccttctatcatg 389

## RESULT 9

V52287

ID V52287 standard; DNA; 1953 BP.

XX AC V52287;

XX AC V52287;

XX DT 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:154.

DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

XX WO9818931-A2.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-US9588.

XX PR 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

PI Kunsch CA, Rosen CA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT

pharmaceutical compositions and vaccines for Streptococcus pneumoniae

Claim 1; Page 1020-1021; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (W521334 to W52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (W521334 to W52524) are genomic fragments from *Streptococcus pneumoniae*. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the *S. pneumoniae* genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridise to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating the nucleic acid molecules from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the *S. pneumoniae* genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the *S. pneumoniae* genome of commercial importance, or expression modulating fragments of the *S. pneumoniae* genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for *S. pneumoniae*.

Sequence 1953 BP; 568 A; 404 C; 443 G; 538 T; 0 other;

Query Match 13.8%; Score 56.4; DB 19; Length 1953;  
Best Local Similarity 48.4%;  
Pred. No. 7.3e-06;  
Matches 193; Conservative 0; Mismatches 197; Indels 9; Gaps 1;

QY	12	goggtcatgggcagaaacctgccctcaacattgcagagaaggggtcccccattctgtg	71
Db	44	gcgctaattgggtcgtaaccttgcccttaattgaatctcgtggttacacagttgctatc	103
QY	72	tacaacagacaacctcaaggtggacgagacogtgcagogtgcgaaggcagagaagaac	131
Db	104	tacaaccgtagtaaaagaaaaacggaagatgtgatgttcgtgcoactcgtgaaagaacttt	163
QY	132	cttcccgctacgggttcacatgcaccccgctcccttttgaagtcacattcagaagccacgg	191
Db	164	gtaccaagct-----atgcgctgaaagttttgaaactcaatcagaaaaacctcgt	214
QY	192	gtgggtatcatgctogtcaaggccggcgccaggttgacacaccatcgcagcgtcgca	251
Db	215	cgtatcatgctgatggttcaagctggacctgggtacagatgctactatccaagcccttctt	274
QY	252	gtcctacttggagcagggcgagctgcatactcgatgggggaaacgagtggtacagagaacag	311
Db	275	ccacaccttgacaaggggtgatattgtatgcagggggaataacttcttcacaaagatacc	334
QY	312	gagaggaggagaagggcatggagagcggcgccctnctgtatcttggcgatgggtgtctct	371
Db	335	atccgctgtaatgaagaattggcaaacctcgttatcaactttatcggctactgggggttct	394
QY	372	ggaggaagaggaggtgcccgcaacggcccgctccttgatg	410
Db	395	ggcgctgaaaaaggtggcccttgagggttcctctcatcatg	433

RESULT 10  
Z30047  
ID Z30047 standard; DNA; 2087 BP.  
XX  
XX  
AC Z30047;  
XX  
XX  
DT 26-JAN-2000 (first entry)  
XX  
XX  
DE DNA encoding *S. pneumoniae* 6-phosphogluconate dehydrogenase protein.

XX KW KW KW XX OS XX FH FT FT FT XX PN XX PD XX PF XX PR XX PA XX PI XX DR DR XX

XX PS XX CC

0000000000000000xxss

Qu  
Be  
Ma

QY Db QY Db QY Db QY

6-phosphogluconate dehydrogenase; bacterial adhesion; antibacterial infection; *Helicobacter pylori*; meningitis; chromosomal mapping; genetic immunisation; ss.

*Streptococcus pneumoniae*.

Key	Location/Qualifiers
CDS	11..1435

```

/product= "6-phosphogluconate dehydrogenase"

```

WO9953020-A1.

21-OCT-1999.

09-APR-1999: 99WO-US07782

10-APR-1998. 98US-0058692

/CMTK \ SMYTHNY TNE PRCNNY CCRS

[illegible]

Warren RL;

WPI; 1999-620409/53.

P-PSDB; Y43524.

New 6-phosphogluconate dehydrogenase from *Streptococcus pneumoniae*, and related proteins, used to screen for antibacterial agents, to treat infections and to raise diagnostic antibodies -

Claim 2; Page 3-4; 62pp; English.

The present sequence encodes a 6-phosphogluconate dehydrogenase protein from *Streptococcus pneumoniae*. The protein induces an immune response against bacteria, causing production of antibodies that interfere with bacterial adhesion. The 6-phosphogluconate dehydrogenase, its fusion proteins and cells (or their membranes) are used to screen for specific modulators of its function or expression, which are potentially useful as antibacterials and to raise antibodies. Antibodies are used as immunonassay reagents; for isolation or identification of 6-phosphogluconate-expressing clones; for affinity purification; and as therapeutic antibacterials. Conditions that can be treated include bacterial infections, by *Helicobacter pylori*, or especially by *S. pneumoniae*, particularly meningitis. Detecting the presence of the protein or mutations in gene sequences encoding it can be used for the staging or monitoring treatment of infections, or determining susceptibility to them, particularly in standard immunoassays or hybridisation and amplification tests. Nucleic acids encoding the protein may also be used for chromosomal mapping and identification, for genetic immunisation or the rational design of (ant)agonists.

Sequence 2087 BP; 610 A; 424 C; 474 G; 572 T; 7 other;

Query Match 13.8%; Score 56.4; DB 20; Length 2087;  
Best Local Similarity 48.4%; Pred. No. 7.5e-06;  
Matches 193; Conservative 0; Mismatches 197; Indels 9; Gaps 1;

QY	12	gcggctatggg	cagaaccttgcctcaacattgcagagaaagggttccccatctctgtg	71
Db	44	gccgtaatggg	cgtaaccttgccttaattgaatcwcgtgggttacacagttgcgtatc	103
QY	72	tacacagggaac	acctccaaggtgcgcagacgcgtgcagcgtgccaaaggcagaaggaac	131
Db	104	tacaacggtag	taaaagaaaaaacggaatggtgattgctgccatctctgaaaaaaccttt	163
QY	132	cttcgcgtact	acgcttccatgaccocgcgtctttgtgaagtcattcagaagccacgg	191
Db	164	gtaccaagct	-----atgcgctgaaagtttgttaaacctcaatgaaaaacctcgt	214
QY	192	gtggtgatcat	gctcgtcaaggccgcgcgcagttgaccagacatctgcgacgtcgca	251

Db 215 cgtatcgtcgtggttcaagctggacctgggtacagatgctactatccaagcccttctt 274

Qy 252 gctcacttgagcagcgactgcacatcgatggggggaacgagtggtacgagaaacag 311  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 275 ccacacttgacaaggtgatattgattgacgwggaataacttctcacaagatacc 334

Qy 312 gagggaggagagagccatggagcggcgccctnctgtattgttgcatgggtgctct 371  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 335 atccgtcgtaatgaagaattggcaaacctcwggtatcaactttatcgtactgrrgttct 394

Qy 372 ggaggaagagaggtgcccgcaacgcccgtccttgatg 410  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 395 ggtgtgaaaagagtgcccttgaaaggtccttctatcatg 433

RESULT 11

V74364/c

ID V74364 standard; DNA; 16592 BP.

XX XX

AC V74364;

XX XX

DT 16-MAR-1999 (first entry)

XX XX

DE Staphylococcus aureus contig SEQ ID #53.

XX XX

KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 skin infection; surgical wound infection; scalded skin syndrome;  
 toxic shock syndrome; ds.

XX OS Staphylococcus aureus.

XX XX

PH Key Location/Qualifiers

FT misc\_feature 121..180

FT /\*tag= a

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 1921..1980

FT /\*tag= b

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 3721..3780

FT /\*tag= c

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 5521..5580

FT /\*tag= d

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 7321..7380

FT /\*tag= e

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 9121..9180

FT /\*tag= f

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 10921..10980

FT /\*tag= g

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 12721..12780

FT /\*tag= h

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 14521..14580

FT /\*tag= i

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 16321..16380

FT /\*tag= j

FT /\*note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

XX XX

PN EP786519-A2.

XX XX

PD 30-JUL-1997.

XX XX

PF 07-JAN-1997; 97EP-0100117.

XX XX

PR 05-JAN-1996; 96US-0009861.

XX XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 Rosen CA;

PI PI

XX WPI; 1997-374922/35.

DR

XX Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 stored on computer readable medium and used in the production of  
 anti-S.aureus vaccines

XX Claim 1; Page 403-413; 327lpp; English.

XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 of the invention. The DNA sequences are recorded on a computer readable  
 medium, preferably selected from a floppy or hard disk, random access  
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 the S.aureus DNA sequences allows putative functions to be assigned so  
 that protein-encoding or regulatory regions of commercial; therapeutic or  
 industrial importance can be obtained. Specifically, sequences which are  
 likely to encode antigens have been identified and these polypeptides can  
 be used in a vaccine composition against S.aureus infection. The  
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 skin and surgical wound infections, scalded skin syndrome, toxic shock  
 syndrome, etc. Organisms transformed with the DNA sequences can be used  
 for recombinant production of the polypeptides. The new DNA sequences  
 (and their fragments) are useful as primers or probes for isolating  
 homologues of any of the S.aureus DNA sequences contained on the  
 computer readable medium.

XX SQ Sequence 16592 BP; 5028 A; 3005 C; 2294 G; 5653 T; 612 other;

Query Match 12.1%; Score 49.6; DB 18; Length 16592;  
 Best Local Similarity 47.1%; Pred. No. 0.00075;  
 Matches 188; Conservative 1; Mismatches 201; Indels 9; Gaps 1;

Qy 12 gcggtcattggggagagacccttgccctcaacattgcagagaaagggtcccatctctgtg 71  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 7899 GCCGTTATGGTAAAAACCTGCTTGGAAATATTGAATCAGCGGATATAGTGTATCTGTA 7840

Qy 72 tacaacaggacaacctccaagggtggacgagaccgtgcagcgtgccaggcagaagaaac 131

Db 7839 TTTAACCGCTCAAGTGAATAAACTGATTTAATGGTTGAAGATCAAAAGGAAAAATATT 7780  
 QY 132 cttcccgctacaggtccatgaccccgctcctttgtgaagtcattcagaagccacgg 191  
 Db 7779 CATCCAACGATATTTCATTGGAAGA-----ATTGTGTAATCTTTAGAAAACACCGT 7729  
 QY 192 gttgtgatcgtcgtcaaggccgcccaggttgaccagaccatcgacgactcgca 251  
 Db 7728 AAAATTTTATTAATGTTTCAAGCAGGTAAAGCTACGGACGCAACGANTGATAGTTGTTA 7669  
 QY 252 gctcattggagcaggcgactgcacatcgatggggggaacgagtggtacgagaacacg 311  
 Db 7668 CCTTATTAGATGATGGAGATATTTTAAATGATGGTATATACAAACTATCAAGATACT 7609  
 QY 312 gagaagagggaagccatgagagcgccgcccctctgtatcttgcgagtggtgtctct 371  
 Db 7608 ATCAGACGTAATAAGCAATTWRCAAAAGTGCATCAACTTATTTGTTATGGCGGTTTCT 7549  
 QY 372 ggaggaaggaggtgcccgcacacgcccgcctcctgatg 410  
 Db 7548 GGTGCTGAATGGAGCGTTAACTGGTCCGTCATTAAAG 7510

## RESULT 12

V21187/c  
 ID V21187 standard; DNA; 53789 BP.

XX AC V21187;

XX XX

DT 24-JUL-1998 (first entry)

XX XX

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

XX XX

KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

XX polyketide synthase; actinomycete; ansamycin; ds.

XX OS Amycolatopsis mediterranei.

XX XX

FT Key Location/Qualifiers

FT CDS 1825..15543

FT /\*tag= a

FT /label= ORF\_A

FT /product= "polyketide synthase"

FT 15550..30759

FT /\*tag= b

FT /label= ORF\_B

FT /product= "polyketide synthase"

FT 30895..36060

FT /\*tag= c

FT /label= ORF\_C

FT /product= "polyketide synthase"

FT 36259..41325

FT /\*tag= d

FT /label= ORF\_D

FT /product= "polyketide synthase"

FT 41373..51614

FT /\*tag= e

FT /label= ORF\_E

FT /product= "polyketide synthase"

FT 51713..5293

FT /\*tag= f

FT /label= ORF\_F

FT /product= "polyketide synthase"

FT XX

PN WO9807868-A1.

XX XX

PD 26-FEB-1998.

XX XX

PF 18-AUG-1997; 97WO-EP04495.

XX XX

PR 20-AUG-1996; 96EP-0810551.

XX XX

(NOVS ) NOVARTIS AG.

XX Engel N, Schupp T, Toupet C;

XX WPI; 1998-169172/15.

DR P-PSDB; W52845-W52850.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used

XX to produce rifamycin and rifamycin analogues

XX Claim 4; Page 53-102; 205pp; English.

XX The present sequence represents a Amycolatopsis mediterranei rifamycin

XX synthesis gene cluster DNA fragment from the present invention. The

XX DNA fragment comprises a DNA region involved directly or indirectly

XX in the gene cluster responsible for rifamycin synthesis, including

XX the adjacent DNA regions to the right and left which, by reason of

XX their function in connection with rifamycin biosynthesis, qualify

XX as constituents of this rifamycin gene cluster, and functional

XX fragments, derivatives or constituents of these. The Amycolatopsis

XX mediterranei rifamycin synthesis gene cluster DNA fragment can be used

XX for producing rifamycin, rifamycin analogues or precursors. It can also

XX be used for inactivating or modifying genes involved in ansamycin or

XX rifamycin biosynthesis. The DNA can be used for constructing mutant

XX actinomycetes strains from which the natural rifamycin or ansamycin

XX biosynthesis gene cluster has been partly or completely deleted. The

XX DNA fragment can be used for assembling a library of polyketide

XX synthases, which can be used for assembling a library of polyketides.

XX A hybridisation probe of the invention can be used for identifying DNA

XX fragments involved in the biosynthesis of ansamycins.

XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

XX

SQ

Query Match 9.6%; Score 39.2; DB 19; Length 53789;

Best Local Similarity 45.2%; Pred. No. 0.52;

Matches 183; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

QY 1 ccaacgctcccggtcgtatgggggagaaacacttgcctcaacattcagagaaggttcc 60

Db 21394 CCGGCGCTCTCGGCGCATGCGCGTCCGCGACATGCGGATGACACCGATCGCGGAGC 21335

QY 61 ccattctgtgtacacaggaacaaactccaagtgtagacagaccgtgagcggtgcaagg 120

Db 21334 CCATCAGCGCGACACCGCGCGCCAGGCGCCATCGTCACTCGCCCTCGCGAGTGTGTG 21275

QY 121 cagaaggaacacctcccgctctacggtcttcacatgaccccgctccctttgtgaagtcattc 180

Db 21274 CGGCGAGGTGGATCGGACCGACCGAGCAGCAGCGCGCTGCCACCGTACCCCGGGGC 21215

QY 181 agaagccacgggtggtgatcatgctctcaaggccgagccagttgaccagaccatcg 240

Db 21214 CCTCGAACCCGAAAGGTGTAGGACACCGCGCGAGCGAGCGCTGCCCGCTGGCGTGG 21155

QY 241 cgacgtcgagctcactgtgagcagcgagcgactgcattcatcgatgggggagagtggt 300

Db 21154 TGACCAAGCCTTCGAGCTCGG---CGGGCATGTTGTCAGGACTCGAGGTAGTCGTGA 21098

QY 301 acgagacacagagagagggagagggccatggagagcgcgccctnctgtatcttgaca 360

Db 21097 TGGAGACCGCGGAGAAAGACCGCGCGCTCACCGCGGGCCTTGGTCGGGTTCGATGCCCG 21038

QY 361 tgggtgtctctggaggaagaggggtgcccgcacacgcccgtctct 405

Db 21037 CGCGCTCCAGGCGTTCCAGGAGGCGTTCGAGCAGCAGCGCGTGTCT 20993

XX

RESULT 13

T63571

ID T63571 standard; DNA; 1524 BP.

XX XX

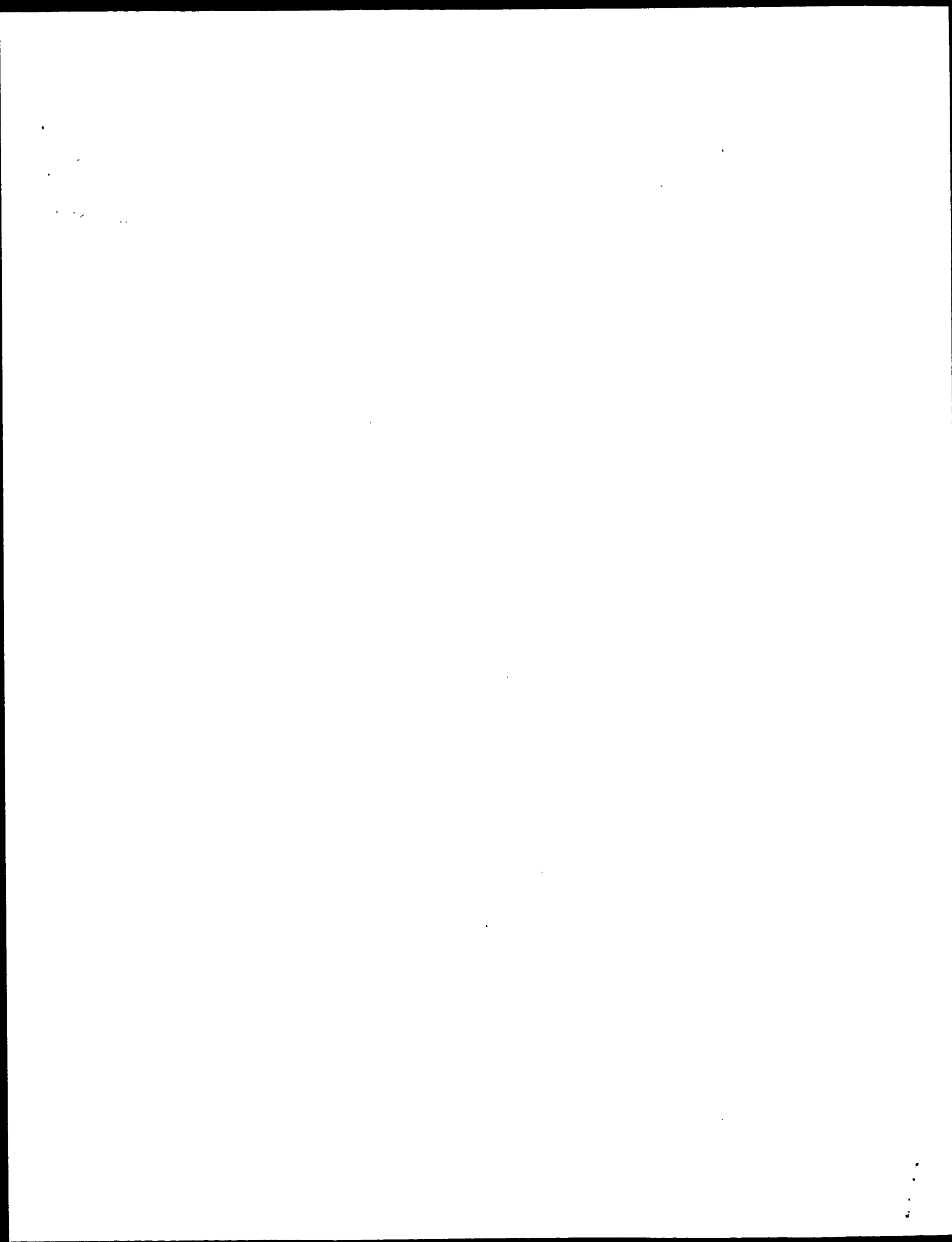
XX T63571;

XX XX











```

; ANTI-SENSE: no
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM:
; FEATURE:
; NAME/KEY: sequence of xylanase gene identified by
; NAME/KEY: amplification of xylanase fragments from soil
US-08-716-942-24

```

```

Query Match          9.1%; Score 37.4; DB 3; Length 1524;
Best Local Similarity 50.6%; Pred. No. 0.2;
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 189 cgggtgtgatcgtctcaaggcgccgagcagtggtgaccagaccatcgagcgtc 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 CGCGGATGGGACGCGGTACATCGCGGAGCGGTGCGCACCGCGCACGCGCGACCC 694

Qy 249 gcagctcaactggagcagggcgactgcatcgatggggggaacgagtggtacgagaac 308
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 AGTGCCAGCTGTACATCAAGACTACAGCATCGAGGCGGAGAACGCCAAGAGCAAGGCC 714

Qy 309 acgagagagggggaagccatgagagcgccgctnctgtatcttggcatggg 364
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 ATGTACAGCTGGTGCAGTCCCTGCTGCGCAGGGGTGCGGATCAACAGCGGTGGG 770

```

```

RESULT 2
US-08-576-626A-2
; Sequence 2, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-576-626A-2

```

```

Query Match          8.7%; Score 35.6; DB 4; Length 8051;
Best Local Similarity 45.6%; Pred. No. 1;
Matches 125; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 71 gtacacagagacaacctccaaggtggacgagaccgtgcagcgtgccaaaggcagagaaa 130
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2317 GTCCGACACACACTCCGGAGGGAGCAGGATGTACAGAGGGGGGTTCGCCGACGTTTA 2376

Qy 131 cettccccgtctacggcttccatgaccccgcgctcttttgaagtccattcagaagccacg 190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2377 CGACCGGTTTACCGCGCGCGGGGCAAGGACTACGCGCGCCGAGCGCGCAGTCCGCGC 2436

Qy 191 ggtgtgatcgtctcgtaaggccgagccgagccaggttgaaccagaccatcgagcgtcgc 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2437 GCTGTGTACAGAGACCCTGCGCTCGGCTTCCTGCTGTCTGCTGCTGCTGCTGCTGCTG 2496

Qy 251 agctcacttggagcagggcgactgcatcgtcgtggtgagcagcagtggtgtacgagaacac 310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2497 CACCCACCTGCGCGGTTGCGCGACCTCTCTGACGACGTGACCGGGCTGGAGCTGTCGCGC 2556

Qy 311 ggagagaggggggagagggccatgagagagcgcgcgc 344
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2557 GGCGATGATCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2590

```

```

RESULT 3
US-08-572-951-1/c
; Sequence 1, Application US/08572951
; Patent No. 5824790
; GENERAL INFORMATION:
; APPLICANT: KEELING, PETER L.
; APPLICANT: KNIGHT, MARY E.
; APPLICANT: GUAN, HANPING
; TITLE OF INVENTION: MODIFICATION OF STARCH
; TITLE OF INVENTION: SYNTHESIS IN PLANTS
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; ADDRESSEE: Intellectual Property Group of
; ADDRESSEE: Pillsbury Madison & Sutro LLP
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/572,951
; FILING DATE: 15-DEC-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/346,602
; FILING DATE: 29-NOV-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/263,921
; FILING DATE: 21-JUN-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul N. Kokulis
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 222957/1.02.15C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```







Sat Nov 4 18:11:10 2000

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/720,229  
FILING DATE: 26-SEP-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Schiff, J. Michael  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20002.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2713 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-720-229-13



```

Db 1108 GAGTTCTTCGATGCGACGACGAGACCTGTCCGACACAGAGAAATGTCCCAAGAGCATC 1167
QY 72 tacaacagga---caacctccaagtggaagagagacggtgcagcgtgcacaagcagaagga 128
Db 1168 ACCAAGTGGAGCTCAATGACCTCATGGACAAGATCGAGAGCCCGAGCCGGAAGACACA 1227
QY 129 aacctcccggtacggtcttcacatgaccccgctcctttgtgaagtccattcagaagcca 188
Db 1228 CAAGATGCTCTACCGCGCAGGTCGCCCTGAGTTCAAGGTGGCTCCAGTGTGGAGCAG 1287
QY 189 cgggtgggtacatctcgtcaagcccgccgagccagtgtagccagacacatgcgacgctc 248
Db 1288 CTGAACATCATAGAGGACGAGGTAGCCAGCGCTGTGCTGACCGCCCTCCAAAGATCCAC 1347
QY 249 gcagctcacttgagcagggcgactgcacatcatcgatgggggaagcagtggtacagagaac 308
Db 1348 GTGCTGCTATGCTGCTGCACGAGGACCACTCTGGACACAGGCGCGGGGACCCCGCAGC 1407
QY 309 acggagagaggagga 322
Db 1408 TCCAAGAAGGGCGA 1421

```

## RESULT 13

```

PCT-US92-05532-1/c
; Sequence 1, Application PC/TUS9205532
; GENERAL INFORMATION:
; APPLICANT: Lichenstein, Henri
; TITLE OF INVENTION: Isolation and
; TITLE OF INVENTION: Characterization of a Novel Protease
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, DS, 1.4 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 6.0.4
; SOFTWARE: Microsoft Word Version 4.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05532
; FILING DATE: 19920629
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1910 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double stranded
; TOPOLOGY: circular
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: DNA sequence encoding
; NAME/KEY: Streptomyces lividans protease x
; OTHER INFORMATION: bp 12 to 1910 comprises
; OTHER INFORMATION: S. lividans DNA sequence, including Protease X.
; OTHER INFORMATION: bp 1 to 11 represents cloning vector DNA sequence.
PCT-US92-05532-1

```

```

Query Match      8.2%; Score 33.8; DB 6; Length 1910;
Best Local Similarity 47.8%; Pred. No. 2;
Matches 98; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 96 gacgagaccgtgcagcgtgcacagggcagaggaacaccttcccgctctacggttccatgac 155
Db 551 GCCAACGCGTGTGCTGCGCGGGTGACCGAGGTTCAGCTGACGTGCGCTTCGTGCGACGAG 492

```

```

QY 156 cccgcgtcctttgtgaagttccattcagaagccaggggtggtgatcatgctcgtcaagggc 215
Db 491 GTCGAGGGCTGTGGCTGTGCACATCCCGAGGACCTGCGGCTGCGCTGCTGCTGCGAGCGC 432
QY 216 ggcgcgcagttgaccagaccatcgcgacgctcgacgtcacctcacttggagcagggcgactgc 275
Db 431 GAACCGGAGTGTGTCATGCTGCCCGCGGACACACCCCGCGGCGCGCGCGCGCTCGTCCGC 372
QY 276 atcatcgatggggggaacgagtggt 300
Db 371 CTCGCCGACCTGCGCGGACGACCGGT 347

```

## RESULT 14

```

US-08-272-882D-1/c
; Sequence 1, Application US/08272882D
; Patent No. 5516685
; GENERAL INFORMATION:
; APPLICANT: Lichenstein, Henri
; TITLE OF INVENTION: Isolation and Characterization of No. 5516685el
; TITLE OF INVENTION: Protease from Streptomyces Lividans
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/272,882D
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy
; REFERENCE/DOCKET NUMBER: A-199A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2261 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-272-882D-1

```

```

Query Match      8.2%; Score 33.8; DB 1; Length 2261;
Best Local Similarity 47.8%; Pred. No. 2.1;
Matches 98; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

```

QY 96 gacgagaccgtgcagcgtgcacagggcagaggaacaccttcccgctctacggttccatgac 155
Db 551 GCCAACGCGCTGTGCTGCGCGGGTGACCGAGGTTCAGCTGACGTGCGCTTCGTGCGACGAG 492
QY 156 cccgcgtcctttgtgaagttccattcagaagccacggtgggtgatcatgctcgtcaagggc 215
Db 491 GTCGAGGGCTGTGCGCTGCACATCCCGAGGACCTGCGGCTGCGGCTCCTGCTGAGGCGC 432
QY 216 ggcgcgcagttgaccagaccatcgcgacgctcgacgtcaccttggagcagggcgactgc 275
Db 431 GAACCGGAGTGTGCTGCTGCGCGGCGGACACACCCCGCGGCGCGCGCTCGTCCGC 372
QY 276 atcatcgatggggggaacgagtggt 300
Db 371 CTCGCCGACCTGCGCGGACGACCGGT 347

```

## RESULT 15

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 8.2%; Score 33.8; DB 1; Length 7218;  
Best Local Similarity 2.5%; Pred. NO. 3;  
Matches 8; Conservative 180; Mismatches 138; Indels 0; Gaps 0;  
QY 75 aacaggacaacccctcaaggtgagacagaccgtgagcgtgccaagggcagaaggaacacctt 134  
Db 1379 RRR 1320  
QY 135 cccgtctacgcttcacagcccgcccttcttgaaagtcattcagaagccacgggtg 194  
Db 1319 RRR 1260  
QY 195 gtgatcgtctgcaagcgccgcagttgaccagaccatcgacgctcgagct 254  
Db 1259 RRR 1200  
QY 255 cacttgagcagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 314  
Db 1199 RRR 1140  
QY 315 aggagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 374  
Db 1139 RRR 1080

QY 375 ggaagaggagggtgcccgaacggccc 400  
Db 1079 RRR 1054  
Search completed: November 4, 2000, 13:33:31  
Job time: 16647 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:43:19 ; Search time 320.8 Seconds  
(without alignments)  
268.163 Million cell updates/sec

Title: US-09-300-482-27  
Perfect score: 229  
Sequence: 1 cagacctatttttctgtc.....cttcaatttatggggaacaa 229

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues  
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	33.6	14.7	1105	19 296449	S. pneumoniae deri
C 2	32.4	14.1	910715	20 X20248	Borrelia burgdorfe
C 3	30.4	13.3	1026	21 251822	Staphylococcus aur
C 4	30.4	13.3	15249	18 V74413	Staphylococcus aur
C 5	29.8	13.0	4434	20 X61678	B. burgdorferi ant
C 6	29.8	13.0	4485	20 X61677	B. burgdorferi ant
C 7	29.8	13.0	116277	20 X20249	Borrelia burgdorfe
C 8	29.6	12.9	1435	17 T43709	Human inhibitor of
C 9	29.6	12.9	2580	18 T70838	Human apoptosis in
C 10	29.6	12.9	2589	18 T61590	Human c-IAP1. Hom
C 11	29.6	12.9	3532	18 T72711	Human inhibitor of
C 12	29.6	12.9	3532	20 Z22143	Human cellular inh

13	29.6	12.9	3732	19 V55040	Human HIAP-2 codin
14	29.4	12.8	388	20 V90010	EST clone C2326.
C 15	29	12.7	72604	20 Z10752	Genomic sequence o
16	28.8	12.6	618	21 A16054	Human colon cancer
17	28.8	12.6	625	20 V89678	EST clone DA505.
18	28.4	12.4	360	18 V77318	Staphylococcus aur
19	28.4	12.4	1530	19 V27310	Arabidopsis leaf a
20	28.4	12.4	1579	17 T44594	AF1q cDNA involved
21	28.4	12.4	1664976	19 V21209	Methanococcus jann
C 22	28	12.2	750	20 X30326	DNA encoding a hum
23	28	12.2	2076	21 Z46833	R. sphaeroides gen
24	28	12.2	2759	9 N80149	Insert of plasmid
C 25	27.8	12.1	1969	20 X60804	Human secreted pro
26	27.6	12.1	995	19 X14581	H. pylori GHP0 167
C 27	27.4	12.0	964	18 T79676	BRCA2 cancer susce
C 28	27.4	12.0	1041	18 T79675	BRCA2 cancer susce
C 29	27.4	12.0	1050	18 T79674	BRCA2 cancer susce
C 30	27.4	12.0	1050	18 T79674	Recombinant papill
31	27.4	12.0	1599	17 T05834	Human papillomavir
C 32	27.4	12.0	8010	17 T13413	Streptococcus pneu
C 33	27.4	12.0	10726	19 V52199	Salmonella enteric
34	27.2	11.9	416	20 X06751	Human cancer-relat
35	27.2	11.9	612	20 V08862	Human secreted pro
36	27.2	11.9	695	19 Z27267	Arabidopsis thalia
C 37	27.2	11.9	1070	11 Q05965	Sequence encoding
38	27.2	11.9	1705	19 Z96331	S. pneumoniae deri
39	27.2	11.9	1705	19 V37382	Streptococcus pneu
C 40	27.2	11.9	3246	20 V08861	Human cancer-relat
C 41	27.2	11.9	3247	20 X78784	Human CtIP cDNA.
C 42	27	11.8	251	16 T22960	Human gene signatu
C 43	27	11.8	756	15 Q73729	Rape abscission/de
44	27	11.8	957	14 Q36645	Soluble human inte
45	27	11.8	1008	17 T16482	Soluble interleuki

## ALIGNMENTS

RESULT 1  
Z96449/c  
ID Z96449 standard; DNA; 1105 BP.  
XX AC Z96449;  
XX DT 10-APR-2000 (first entry)  
XX DE S. pneumoniae derived DNA from ORF #277.  
XX KW Treatment; prevention; disease; diagnosis; screening;  
XX KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.  
XX OS Streptococcus pneumoniae.  
XX PN W09806734-A1.  
XX PD 19-FEB-1998.  
XX PF 15-AUG-1997; 97WO-US14436.  
XX PR 16-AUG-1996; 96US-0024022.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
XX Stodola RK;  
XX WPI; 1998-159452/14.  
XX P-PSDB; Y86134.  
XX Streptococcus pneumoniae proteins and related DNA - useful for  
XX screening compounds for antibacterial activity  
XX Claim 4; Page 291; 640pp; English.

XX This invention describes novel isolated Streptococcus pneumoniae  
 CC polynucleotides (see Z96173-Z96494) and their encoded proteins (see  
 CC Y85792-Y86182). The DNA, vectors and host cells described in the method  
 CC of the invention are useful for the recombinant expression of the  
 CC polypeptides. The polypeptides are useful for treatment or prevention of  
 CC disease, or diagnosis of disease related to expression or activity of  
 CC such a polypeptide. They can also be used to screen for compounds which  
 CC interact with and inhibit or activate such a polypeptide. The  
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
 CC for inducing an immunological response in a mammal. The antagonists are  
 CC useful to inhibit such bacterial polypeptides. The polypeptides are  
 CC particularly useful to identify antimicrobial compounds and antibiotics.  
 CC They are also useful to determine their role in pathogenesis of  
 CC infection, dysfunction and disease.

XX Sequence 1105 BP; 303 A; 244 C; 193 G; 365 T; 0 other;

Query Match 14.7%; Score 33.6; DB 19; Length 1105;  
 Best Local Similarity 52.1%; Pred. No. 0.17;  
 Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 65 acaagaatagccttgctggtgctggttaaatggcgaatctggcactcaatattgct 124  
 Db 284 ACTGGCTGGGCTTGTGTTAGCGGGATTTTCCTGCTAAACACAGCTCCACAAACACCT 225  
 Qy 125 tgaagaaggcttcccaattccggttaacacggaacatttcccaaggtatttgggocata 184  
 Db 224 GTCRAAGAGATCGAATTTGACAGTGAAGTCGACAGCTGTTTCCCAAGGATTTGGTATCCGA 165  
 Qy 185 agacgaagcaaacaggagaagaaa 208  
 Db 164 AAGGAGAGTGAACAGGAAGAAA 141

## RESULT 2

X20248/c  
 ID X20248 standard; DNA; 910715 BP.

XX AC X20248;

XX DT 04-MAY-1999 (first entry)

XX DE Borrelia burgdorferi polynucleotide sequence #1.

XX KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.

XX OS Borrelia burgdorferi.

XX PN WO9858943-Al.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US12764.

XX PR 03-SEP-1997; 97US-0057483.

XX PR 20-JUN-1997; 97US-0050359.

XX PR 22-JUL-1997; 97US-0053344.

XX PR 22-JUL-1997; 97US-0053377.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (MEDI-) MEDIMUNE INC.

XX PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;

XX PI White OR;

XX DR WPI; 1999-081217/07.

XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 products for the detection, diagnosis, characterisation, prevention

PT and therapy of infections, particularly Lyme disease

XX PS Claim 1; Page 157-671; 1128pp; English.

XX X20248 to X20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.

XX Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

Query Match 14.1%; Score 32.4; DB 20; Length 910715;  
 Best Local Similarity 49.4%; Pred. No. 4.4;

Matches 84; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 57 aaccacacacagaatagccttgctggtgctggttaatggcgaatctggcactca 116  
 Db 171375 AATTTAAACAAAGATAAATCTTGATGTTGCTGTGCTATTGTCGAAAAATAAATG 171316

Qy 117 atattgcttgaagggttcccaattccggttaacacggaacatttcccaaggtatt 176  
 Db 171315 ATATGTTTATAGCATGCTTTAAAGGTTGAAATGAAGATAAACCTGTTTATTTGATT 171256

Qy 177 gggccataagacgaagcaaacaccaggaagaaaccccttccattatgggaaa 226  
 Db 171255 GTGTTATCAATAAATTACATCTGAAATAATAATCTTTCAAGTAATGAAAAA 171206

## RESULT 3

Z51822  
 ID Z51822 standard; DNA; 1026 BP.

XX AC Z51822;

XX DT 04-JUL-2000 (first entry)

XX DE Staphylococcus aureus glycoprotease (gcp) polynucleotide.

XX KW Staphylococcus aureus glycoprotease; gcp; antibody; vaccine formulation;  
 KW Helicobacter pylori infection; gastrointestinal carcinoma; gastric ulcer;  
 KW gastritis; osteomyelitis; septic arthritis; septic thrombophlebitis;  
 KW acute bacterial endocarditis; Staphylococcus food poisoning;  
 KW scalded skin syndrome; toxic shock syndrome; antibacterial; cytostatic;  
 KW antiulcer; antiarthritic; vulnery; immunosuppressive; ss.

XX OS Staphylococcus aureus.

XX FH Key Location/Qualifiers

XX FT CDS 1..1026 a /product= "Glycoprotease polypeptide"

XX PN WO200013694-Al.

XX PD 16-MAR-2000.

XX XX 23-AUG-1999; 99WO-US19273.

XX XX 08-SEP-1998; 98US-0149624.

XX XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX XX Palmer LM, Kosmatka AL, Traini CM, Warren RL;

XX XX WPI; 2000-256855/22.

XX XX P-PSDB; Y70524.



CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the *Borrelia* genus. The products can also  
CC be used for detection of members of the *Borrelia* genus.

XX

SQ Sequence 4434 BP; 1616 A; 416 C; 740 G; 1662 T; 0 other;

Query Match 13.0%; Score 29.8; DB 20; Length 4434;  
Best Local Similarity 56.7%; Pred. No. 4, 6;  
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Db 766 ATTTTCCCTGCAATTATATAAATCTTAATATAAACCACATCA 707

Qy 69 gaataggcctgcgtgattggcgtgttaatgggcacaaa 105

Db 706 CAAATACCTCCCTTACGATTTGCTATCTAATAAAGAGAAA 670

KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.  
XX  
XX Borrelia burgdorferi.  
OS  
XX  
XX  
XX W00950071-A1  
DN

PR 22-JUL-1997; 97US-0053377.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMMUNE INC.

XX  
PS Claim 1; Page 146-147; 275pp; English.  
XX  
CC This sequence encodes a *Borrelia burgdorferi* (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC

Query Match 13.0%; Score 29.8; DB 20; Length 4485;  
Best Local Similarity 56.7%; Pred. No. 4.7;  
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0



Db 1298 aggtactgttcgtacattctctctaaagaaataagctctatttttaacctgcataa 1357  
 QY 128 aaaggcttcccaattccggttaacaacggaaccatttccaaaggtatttgggcataa 187  
 Db 1358 aaaggtctttaaattgttgaaacacttgaagccatctaaagtaaaagggaattatga 1417

## RESULT 9

T70838  
 ID T70838 standard; cDNA; 2580 BP.

XX T70838;

XX 02-SEP-1997 (first entry)

XX Human apoptosis inhibitor hiap-2 cDNA.

XX Apoptosis inhibitor; HIAP-2; HIV; AIDS; neurodegeneration;  
 KW myelodysplastic syndrome; ischaemia; myocardial infarction; stroke;  
 KW reperfusion injury; toxin-induced liver disease; gene therapy;  
 KW diagnosis; ds.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH 238..2094  
 FT CDS /\*tag= a

XX PN WO9706255-A2.

XX PD 20-FEB-1997.

XX PF 05-AUG-1996; 96WO-IB01022.

XX PR 22-DEC-1995; 95US-0576956.

XX PR 04-AUG-1995; 95US-0511485.

XX PA (UYOT-) UNIV OTTAWA.

XX Baird S, Korneluk RG, Liston P, Mackenzie AB;

XX WPI: 1997-154262/14.

XX P-PSDB; W19583.

XX Nucleic acid encoding an inhibitor of apoptosis polypeptide - used  
 PT to inhibit apoptosis in e.g. HIV or AIDS patients, and for detection  
 PT of susceptibility to apoptotic disease

XX Claim 12; Page 74-75; 219pp; English.

XX Human xiap, hiap-1 and hiap-2 genes, and murine xiap, hiap-1 and  
 CC hiap-2 genes (T70836-41) respectively code for a new class of  
 CC mammalian proteins (W19581-86) that are inhibitors of apoptosis  
 CC (IAP). The hiap-1 and hiap-2 genes were cloned by screening a  
 CC human liver library with a probe including the entire xiap coding  
 CC region. They were assigned to chromosome 11q22-23 by FISH. IAP  
 CC nucleic acids can be used to express IAP polypeptides in cells and  
 CC animals to inhibit apoptosis, and as primers and probes to identify  
 CC and isolate additional IAP genes, as well as in methods for treating  
 CC diseases and disorders involving apoptosis (anti-apoptotic gene  
 CC therapy).

XX Sequence 2580 BP; 800 A; 445 C; 534 G; 799 T; 2 other;

Query Match 12.9%; Score 29.6; DB 18; Length 2580;  
 Best Local Similarity 47.8%; Pred. No. 4.5;  
 Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 8 tattttttgtcattgtctcaatttcaggagagattattatgcgtcaaccacaaca 67  
 Db 2006 tagtatgccagggaatgtgccctctctctaaagaaatgcctatttgcagggtataatca 2065

QY 68 agaataggccttgcgtgattggctgttaattgggcaaatctggcactcaattgtctga 127  
 Db 2066 aggtactgttcgtacattctctctctaaagaaataagctctatttttaacctgcataa 2125  
 QY 128 aaagggtctcccaattccggttaacaacggaaccatttccaaaggtatttgggcataa 187  
 Db 2126 aaaggtctttaaattgttgaaacacttgaagccatctaaagtaaaagggaattatga 2185

## RESULT 10

T61590

ID T61590 standard; cDNA; 2589 BP.

XX T61590;

XX 22-JUL-1997 (first entry)

XX Human c-IAP1.

XX IAP; inhibitor; apoptosis; RING finger domain; restinosis;  
 KW myocardial infarction; nephritis; HIV; ss.

XX Homo sapiens.

XX WO9706182-A1.

XX PD 20-FEB-1997.

XX PF 06-AUG-1996; 96WO-US12860.

XX PR 08-DEC-1995; 95US-0569749.

XX PR 08-AUG-1995; 95US-0512946.

XX PA (TULA-) TULARIK INC.

XX Goeddel DV, Rothe M;

XX WPI: 1997-154209/14.

XX P-PSDB; W13545.

XX Nucleic acids encoding cellular inhibitor of apoptosis proteins -  
 PT useful for apoptosis regulation in cells to reduce or increase  
 PT apoptosis and for pharmacological screening

XX Claim 1; Page 16-18; 35pp; English.

XX The human cellular inhibitor of apoptosis proteins (c-IAP1/2 -  
 CC T61590/T61591) comprise a series of defined structural domain  
 CC repeats and/or a RING finger domain; in particular, at least two of  
 CC a first domain repeat (W13547 or W13548), a second domain repeat  
 CC (W13549 or W13550), and a third domain repeat (W13551 or W13552)  
 CC and/or a RING finger domain (W13553 or W13554), or a consensus  
 CC sequences derived from these human genes.

XX The nucleic acid is used for recombinant prodn. of human cellular  
 CC inhibitor of apoptosis protein which modulates apoptosis  
 CC regulation. The nucleic acids are useful in therapies where  
 CC increased cell-specific apoptosis is desired, e.g. in restinosis,  
 CC inflammatory disease states, myocardial infarction, glomerular  
 CC nephritis, transplant rejection and infectious diseases, e.g. HIV.  
 CC They can also be used in conditions requiring a reduction in  
 CC apoptosis.

XX Sequence 2589 BP; 844 A; 436 C; 522 G; 787 T; 0 other;

Query Match 12.9%; Score 29.6; DB 18; Length 2589;  
 Best Local Similarity 47.8%; Pred. No. 4.5;  
 Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 8 tattttttgtcattgtctcaatttcaggagagattattatgcgtcaaccacaaca 67  
 Db 1963 tagtatgccagggaatgtgccctctctctaaagaaatgcctatttgcagggtataatca 2022

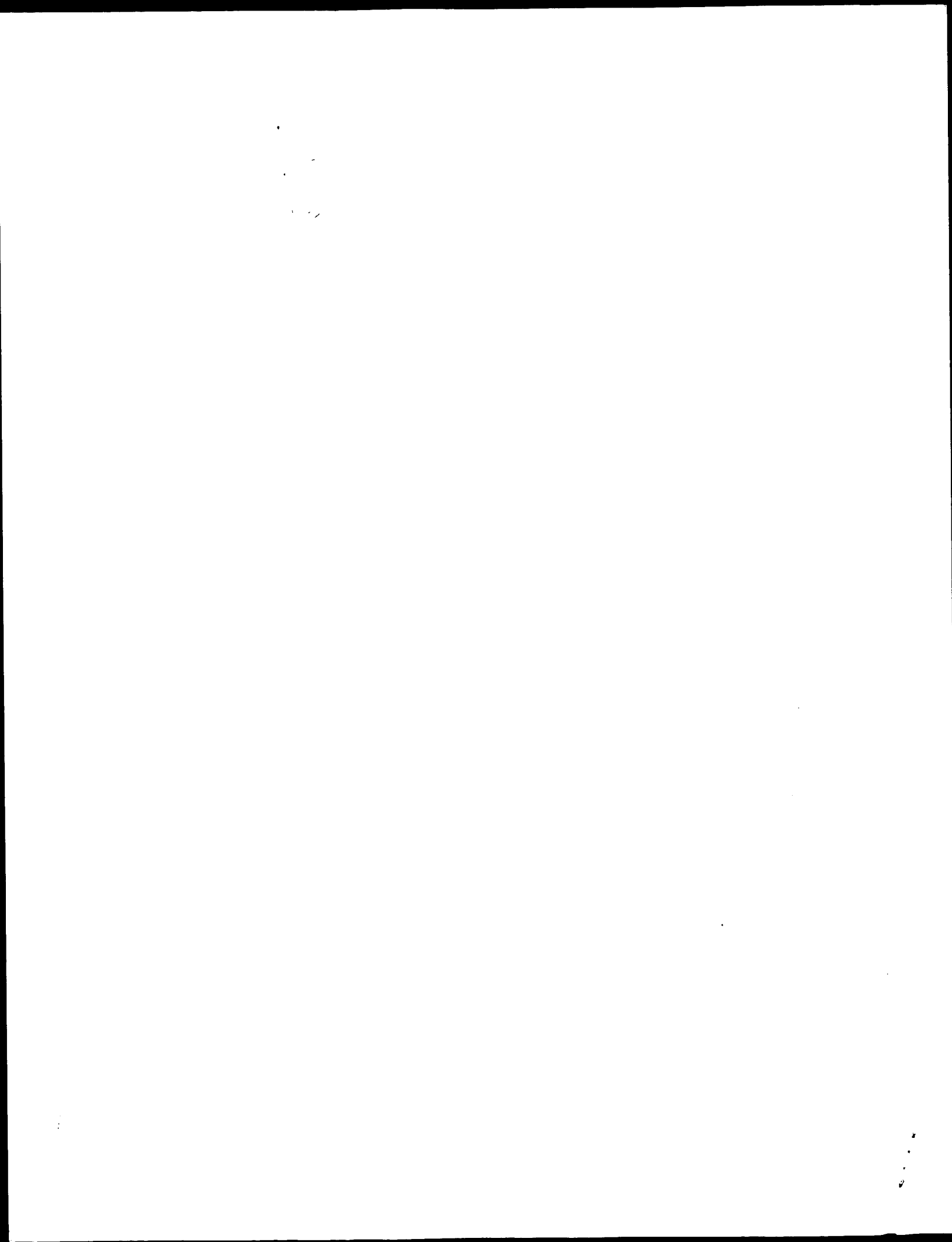


[illegible]

8	tattttttctgctatttgccttcaaatctcaggagattaaattacgcgtcaacccaaca	67
bb		
2928	tagcatgcaggaaatgcacctctctaaagaaatccctatttcagggggtaatca	2987
68	agaataggcccttgctgatcggctgttaatggcgaatactggcaactcaatatgcttga	137
2988	aggactgtctgcacattctctctaaagaaaaaataagctctatttttaacctgctaa	3047
bb		







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:33:31 ; Search time 189.35 Seconds  
(without alignments)  
182.912 Million cell updates/sec

Title: US-09-300-482-27

Perfect score: 229

Sequence: 1 cagaccttatttttctgtc.....cttcaatttatgggaacaa 229

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29.6	12.9	1435	6	PCT-US95-05922A-1
2	29.6	12.9	2580	4	US-08-511-485-7
3	29.6	12.9	2589	6	PCT-US96-12860-1
4	29.6	12.9	3532	4	US-09-205-204-1
5	29.6	12.9	3732	5	US-09-212-971-7
6	28.4	12.4	270	2	US-08-589-080-3
7	28.4	12.4	1629	2	US-08-589-080-2
8	27.4	12.0	964	5	US-08-755-587-23
9	27.4	12.0	1041	5	US-08-755-587-22
10	27.4	12.0	1050	5	US-08-755-587-21
11	27.4	12.0	1599	5	US-08-737-336-5
12	27.4	12.0	8010	6	PCT-US95-11859-2
13	27	11.8	756	3	US-08-530-165-1
14	27	11.8	1008	1	US-07-904-073-1
15	27	11.8	1008	1	US-07-904-071-1
16	27	11.8	1770	3	US-08-381-603-1
17	27	11.8	1770	6	PCT-US94-02414-1
18	27	11.8	1770	6	PCT-US96-08899-1
19	27	11.8	3011	1	US-07-821-716-1
20	26.6	11.6	714	5	US-08-643-704A-44
21	26.6	11.6	816	5	US-08-643-704A-47
22	26.6	11.6	900	5	US-08-643-704A-46
23	26.6	11.6	1236	5	US-08-643-704A-48
24	26.6	11.6	4649	7	5183745-1
25	26.6	11.6	6443	7	5183745-5
26	26.6	11.6	9432	1	US-08-277-231A-1
Sequence 1, Appli					

C 27	26.6	11.6	9432	3	US-08-473-750-4	Sequence 4, Appli
C 28	26.6	11.6	9432	4	US-08-477-326-4	Sequence 4, Appli
C 29	26.6	11.6	17041	1	US-08-076-011-1	Sequence 1, Appli
C 30	26.2	11.4	1194	5	US-08-929-329-3	Sequence 3, Appli
C 31	26	11.4	2095	2	US-08-405-230-8	Sequence 8, Appli
C 32	26	11.4	2095	3	US-08-910-990-8	Sequence 8, Appli
C 33	26	11.4	2564	5	US-08-276-968A-19	Sequence 19, Appli
C 34	25.8	11.3	6973	2	US-08-478-370-1	Sequence 1, Appli
C 35	25.6	11.2	816	5	US-08-968-563-8	Sequence 8, Appli
C 36	25.4	11.1	722	2	US-08-698-551-5	Sequence 5, Appli
C 37	25.4	11.1	722	3	US-08-602-228-5	Sequence 5, Appli
C 38	25.4	11.1	722	3	US-08-649-341A-5	Sequence 5, Appli
C 39	25.4	11.1	722	3	US-08-494-440B-5	Sequence 5, Appli
C 40	25.4	11.1	722	3	US-08-533-901B-5	Sequence 5, Appli
C 41	25.4	11.1	722	3	US-08-839-032A-5	Sequence 5, Appli
C 42	25.4	11.1	722	4	US-08-839-031A-5	Sequence 5, Appli
C 43	25.4	11.1	722	6	PCT-US95-12724-5	Sequence 5, Appli
C 44	25.4	11.1	1018	1	US-08-181-271A-30	Sequence 30, Appli
C 45	25.4	11.1	1018	1	US-08-449-315-30	Sequence 30, Appli

## ALIGNMENTS

RESULT 1  
PCT-US95-05922A-1  
; Sequence 1, Application PC/TUS9505922A  
; GENERAL INFORMATION:  
; APPLICANT: HE, ET AL.  
; TITLE OF INVENTION: Human Inhibitor of Apoptosis Gene 1  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/05922A  
; FILING DATE: 11 MAY 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-292  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1435 BASE PAIRS  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: CDNA  
PCT-US95-05922A-1

Query Match 12.9%; Score 29.6; DB 6; Length 1435;  
Best Local Similarity 47.8%; Pred. No. 0.64;  
Matches 86; Conservative 94; Indels 0; Gaps 0;

QY 8 tatttttgcattgtctcaatttcaggagattaattgctcaccaccaaca 67

Sat Nov 4 18:11:18 2000

Db 1238 TAGTATGCCAGGAATGCCCCCTTCTCTAAGAAATGCCCTATTTCAGGGGTATAATCA 1297  
Qy 68 aagaatagcccttgctggtatggttgaatggcacaatctggcaactcaatattgcttga 127  
Db 1298 AGGTACTGTCGACATTTCTCTTAAGAAATAGTCTATATTTAACTGCATAA 1357  
Qy 128 aaagggtcccaattccggttaacacggaacatttcccaaggttattggccataaga 187  
Db 1358 AAAGGTCTTTAAATATTGTTGAACACTTGAAGCCACTAAAGTAAAGGAATATGA 1417

## RESULT 2

US-08-511-485-7  
; Sequence 7, Application US/08511485  
; Patent No. 5919912  
; GENERAL INFORMATION:  
; APPLICANT: Korneluk, Robert G.  
; APPLICANT: Mackenzie, Alexander E.  
; APPLICANT: Baird, Stephen  
; TITLE OF INVENTION: MAMMALIAN IAP GENE FAMILY, PRIMERS,  
; TITLE OF INVENTION: PROBES, AND DETECTION METHODS  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/511,485  
; FILING DATE: 04-AUG-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 07540/002001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2580 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: DNA (genomic)  
US-08-511-485-7

Query Match 12.9%; Score 29.6; DB 4; Length 2580;  
Best Local Similarity 47.8%; Pred. No. 0.84; Indels 0; Gaps 0;  
Matches 86; Conservative 0; Mismatches 94;  
Qy 8 tatttttctgcttgcattgcttcaaatctcaggagatttaattatgcgctcaacccacaaca 67  
Db 2006 TAGTATGCCAGGAATGCCCCCTTCTCTAAGAAATGCCCTATTTCAGGGGTATAATCA 2065  
Qy 68 aagaatagcccttgctggtatggttgaatggcacaatctggcaactcaatattgcttga 127  
Db 2066 AGGTACTGTCGACATTTCTCTTAAGAAATAGTCTATATTTAACTGCATAA 2125  
Qy 128 aaagggtcccaattccggttaacacggaacatttcccaaggttattggccataaga 187  
Db 2126 AAAGGTCTTTAAATATTGTTGAACACTTGAAGCCACTAAAGTAAAGGAATATGA 2185

## RESULT 3

PCT-US96-12860-1  
; Sequence 1, Application PC/TUS9612860  
; GENERAL INFORMATION:  
; APPLICANT: TULARIK, INC.  
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/12860  
; FILING DATE: 06 AUG 1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. Serial Nos. 08/512,946 & 08/569,749  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brezner, David J.  
; REGISTRATION NUMBER: 24,774  
; REFERENCE/DOCKET NUMBER: A-62464/DJB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415)781-1989  
; TELEFAX: (415)398-3249  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2589 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
PCT-US96-12860-1

Query Match 12.9%; Score 29.6; DB 6; Length 2589;  
Best Local Similarity 47.8%; Pred. No. 0.84; Indels 0; Gaps 0;  
Matches 86; Conservative 0; Mismatches 94;  
Qy 8 tatttttctgcttgcattgcttcaaatctcaggagatttaattatgcgctcaacccacaaca 67  
Db 1963 TAGTATGCCAGGAATGCCCCCTTCTCTAAGAAATGCCCTATTTCAGGGGTATAATCA 2022  
Qy 68 aagaatagcccttgctggtatggttgaatggcacaatctggcaactcaatattgcttga 127  
Db 2023 AGGTACTGTCGACATTTCTCTTAAGAAATAGTCTATATTTAACTGCATAA 2082  
Qy 128 aaagggtcccaattccggttaacacggaacatttcccaaggttattggccataaga 187  
Db 2083 AAAGGTCTTTAAATATTGTTGAACACTTGAAGCCACTAAAGTAAAGGAATATGA 2142

## RESULT 4

US-09-205-204-1  
; Sequence 1, Application US/09205204  
; Patent No. 5958772  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; APPLICANT: Elizabeth J. Ackermann  
; APPLICANT: Lex W. Cowser  
; TITLE OF INVENTION: ANTISENSE MODULATION OF CELLULAR INHIBITOR OF APOPTOSIS-1 EXPR  
; FILE REFERENCE: RFS-0020  
; CURRENT APPLICATION NUMBER: US/09/205,204  
; CURRENT FILING DATE: 1998-12-03

NUMBER OF SEQ ID NOS: 47  
SEQ ID NO 1  
LENGTH: 3532  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1160)..(3016)  
US-09-205-204-1

Query Match 12.9%; Score 29.6; DB 4; Length 3532;  
Best Local Similarity 47.8%; Pred. No. 0.97;  
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
QY 8 tattttttctgcttgccttccaaatttcaggagatttaattatgctcaccacacaaca 67  
Db 2928 tagtatgcaggaatgtgcccctctcctaagaataatgcttattgaggggtataatca 2987  
QY 68 agaataggccttgcctgattgctgtaattgggcaaaatctggcactcaaatattgcttga 127  
Db 2988 agggtaactgttcattctctcttcttaagaaaaatagttctatttttaacctgcataa 3047  
QY 128 aaagggtcccaattccggttaacaacgcaaccatttcccaaggttatttgggccataaga 187  
Db 3048 aaaggtctttaaataattgttgaaacacttgagccatctaagtaaaaaagggaattatga 3107

RESULT 5  
US-09-212-971-7  
Sequence 7, Application US/09212971B  
Patent No. 6107041  
GENERAL INFORMATION:  
APPLICANT: Korneluk, Robert G  
APPLICANT: MacKenzie, Alexander E  
APPLICANT: Liston, Peter  
APPLICANT: Baird, Stephen  
APPLICANT: Tsang, Benjamin K  
APPLICANT: Pratt, Christine  
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND  
TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE  
TITLE OF INVENTION: DISEASE  
FILE REFERENCE: 07891/009002  
CURRENT APPLICATION NUMBER: US/09/212.971B  
CURRENT FILING DATE: 1998-12-16  
EARLIER APPLICATION NUMBER: 60/017,354  
EARLIER FILING DATE: 1996-04-26  
EARLIER APPLICATION NUMBER: 60/030,590  
EARLIER FILING DATE: 1996-11-14  
EARLIER APPLICATION NUMBER: 08/800,929  
EARLIER FILING DATE: 1997-02-13  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 3732  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-212-971-7

Query Match 12.9%; Score 29.6; DB 5; Length 3732;  
Best Local Similarity 47.8%; Pred. No. 0.99;  
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
QY 8 tattttttctgcttgccttccaaatttcaggagatttaattatgctcaccacacaaca 67  
Db 3148 tagtatgcaggaatgtgcccctctcctaagaataatgcttattgaggggtataatca 3207  
QY 68 agaataggccttgcctgattgctgtaattgggcaaaatctggcactcaaatattgcttga 127  
Db 3208 agggtaactgttcattctctcttcttaagaaaaatagttctatttttaacctgcataa 3267  
QY 128 aaagggtcccaattccggttaacaacggaaccatttcccaaggttatttgggccataaga 187

Db 3268 aaaggtctttaaataattgttgaaacacttgagccatctaagtaaaaaagggaattatga 3327

## RESULT 6

US-08-589-080-3  
Sequence 3, Application US/08589080  
Patent No. 5808026  
GENERAL INFORMATION:  
APPLICANT: COHEN, Amos  
APPLICANT: TSE, William  
APPLICANT: ZHU, Weimin  
TITLE OF INVENTION: SEQUENCE OF AFIq CDNA  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/589,080  
FILING DATE: 23-JAN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9501386.8  
FILING DATE: 23-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Rea, Teresa Stanek  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 024916-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-589-080-3

Query Match 12.4%; Score 28.4; DB 2; Length 270;  
Best Local Similarity 62.9%; Pred. No. 0.76;  
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 146 ggtaaacacgaacacatttccaaaggttatttggccataagacgaacacacgaagg 205  
Db 117 GGTCAAAGACAGCAGCGTTGGCAAAATGATCGGGCAAGCAACTGCAGCAGCAGGAGAA 176  
QY 206 aaaccttcaa 215  
Db 177 AAACCTGAA 186

## RESULT 7

US-08-589-080-2  
Sequence 2, Application US/08589080  
Patent No. 5808026  
GENERAL INFORMATION:  
APPLICANT: COHEN, Amos  
APPLICANT: TSE, William  
APPLICANT: ZHU, Weimin  
TITLE OF INVENTION: SEQUENCE OF AFIq CDNA  
NUMBER OF SEQUENCES: 6

RESULT 8  
US-08-755-587-23/c  
; Sequence 23, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard F  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer  
; TITLE OF INVENTION: susceptibility gene and uses thereof.  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell Seltzer park & Gibson  
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0. Version #1.25 (EPO)

RESULT 9  
US-08-755-587-22/c  
; Sequence 22, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard F  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer  
; TITLE OF INVENTION: susceptibility gene and uses thereof.  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell Seltzer Park & Gibson  
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/755,587  
; FILING DATE: 25-NOV-1996  
; PRIOR APPLICATION DATA:



```

; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 503..541
; FEATURE:
; NAME/KEY: exon
; LOCATION: 501..541
US-08-755-587-22

```

```

Query Match      12.0%; Score 27.4; DB 5; Length 1041;
Best Local Similarity 53.2%; Pred. No. 3.1;
Matches 58; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 2 agacccatttttctgtcatttgccttcaaaatttcaggagatttaattatgcgtcaacc 61
   |||||
Db 742 ATAGTTTATTTTATAGAAATGCCCTGATCATTAAGGATATATTTTTCATTCACTCT 683
   |||||

Qy 62 acaacaagaataggccttgcgtggtggtggttaattgggcaaaatctgg 110
   |||||
Db 682 TTATTTCACTTAACGTTTACTGATTCGCTGATGAGGAGAGATGCTAG 634
   |||||

```

```

RESULT 10
US-08-755-587-21/c
; Sequence 21, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0

```

```

; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1050 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 502..550
; FEATURE:
; NAME/KEY: exon
; LOCATION: 501..550
US-08-755-587-21

```

```

Query Match      12.0%; Score 27.4; DB 5; Length 1050;
Best Local Similarity 53.2%; Pred. No. 3.1;
Matches 58; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 2 agacccatttttctgtcatttgccttcaaaatttcaggagatttaattatgcgtcaacc 61
   |||||
Db 883 ATAGTTTATTTTATAGAAATGCCCTGATCATTAAGGATATATTTTTCATTCACTCT 824
   |||||

Qy 62 acaacaagaataggccttgcgtggtggtggttaattgggcaaaatctgg 110
   |||||
Db 823 TTATTTCACTTAACGTTTACTGATTCGCTGATGAGGAGAGATGCTAG 775
   |||||

```

```

RESULT 11
US-08-737-336-5
; Sequence 5, Application US/08737336
; Patent No. 6013262
; GENERAL INFORMATION:
; APPLICANT: FRAZER, Ian
; APPLICANT: ZHOU, Jian
; TITLE OF INVENTION: RECOMBINANT PAPILLOMA VIRUS L1
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,336
; FILING DATE: 16-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/AU95/00292
; FILING DATE: 17-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM 5667
; FILING DATE: 17-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 65064/118
; TELECOMMUNICATION INFORMATION:

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:05 ; Search time 4352.3 Seconds  
(without alignments)  
325.314 Million cell updates/sec

Title: US-09-300-482-27

Perfect score: 229

Sequence: 1 cagacccattttttctgtc.....cttcaatttatggggaacaa 229

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estom:\*  
68: em\_estov1:\*  
69: em\_estov2:\*  
70: em\_estpl1:\*  
71: em\_estpl2:\*  
72: em\_estpl3:\*  
73: em\_estpl4:\*  
74: em\_estpl5:\*  
75: em\_estrol:\*  
76: em\_estrol2:\*  
77: em\_estrol3:\*  
78: em\_estrol4:\*  
79: em\_estrol5:\*  
80: em\_estrol6:\*  
81: em\_estrol7:\*  
82: em\_estrol8:\*  
83: em\_estrol9:\*  
84: em\_estrol10:\*  
85: em\_estrol11:\*  
86: em\_estrol12:\*  
87: em\_estrol13:\*  
88: gb\_gss1:\*  
89: gb\_gss2:\*  
90: gb\_gss3:\*  
91: gb\_gss4:\*  
92: em\_gss1:\*  
93: em\_gss2:\*  
94: em\_gss3:\*  
95: em\_gss4:\*  
96: gb\_gss5:\*  
97: gb\_gss6:\*  
98: gb\_gss7:\*  
99: gb\_gss8:\*  
100: gb\_gss9:\*  
101: em\_gss5:\*  
102: em\_gss6:\*  
103: em\_gss7:\*  
104: em\_gss8:\*  
105: em\_gss9:\*  
106: em\_gss10:\*  
107: em\_gss11:\*  
108: gb\_gss10:\*  
109: gb\_gss11:\*  
110: em\_gss12:\*  
111: gb\_gss12:\*  
112: gb\_gss13:\*  
113: gb\_gss14:\*  
114: gb\_gss15:\*  
115: gb\_gss16:\*  
116: gb\_gss17:\*

117: gb\_gss18:\*  
 118: gb\_gss19:\*  
 119: em\_gss13:\*  
 120: gb\_gss20:\*  
 121: gb\_gss21:\*  
 122: gb\_gss22:\*  
 123: gb\_gss23:\*  
 124: gb\_gss24:\*  
 125: em\_gss14:\*  
 126: em\_gss15:\*  
 127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	139.4	60.9	561	32	BE022117 sm67f12.y
2	113.4	49.5	532	24	AW719787 LJNEST10A
3	110.4	48.2	430	18	AV406682 AV406682
4	108.6	47.4	653	24	AW695821 NF099E09S
5	101.4	44.1	499	32	BE020243 sm42d04.y
6	96.2	42.0	573	23	AW586409 EST1318032
7	96.2	41.9	430	14	AL368071 MCB322A07
8	89.6	39.1	420	14	AL362847 MCB310C10
9	87.4	38.2	337	14	AL367204 MCB312H03
10	87.4	38.2	479	14	AL365745 MCB302B10
11	87.4	38.2	481	14	AL365745 MCB302B10
12	87.4	38.2	630	23	AW688619 NF009F08S
13	85.6	37.4	321	18	AV410206 AV410206
14	85.6	37.4	379	18	AV410953 AV410953
15	85.6	37.4	394	18	AV425788 AV425788
16	82.2	35.9	606	21	AW299197
17	80.6	35.2	548	14	AL368600
18	79.4	34.7	733	3	A1055038
19	78.3	34.1	332	34	BE204727
20	78.3	34.1	504	35	BE450814
21	74.6	32.6	361	24	AW760460
22	74.2	32.4	462	39	T04690
23	74.2	32.4	530	19	AV526953
24	72.6	31.7	528	19	AV550009
25	72.6	31.7	528	19	AV550009
26	72.2	31.5	434	21	AW255063
27	71.6	31.3	375	19	AV540033
28	70.8	30.9	498	35	BE341175
29	67.7	29.3	638	21	AW254720
30	66.6	28.8	158	35	BE343110
31	64.6	28.2	592	23	AW560451
32	63.4	27.7	362	35	BE403238
33	63.4	27.7	382	35	BE406322
34	63.4	27.7	401	35	BE428880
35	63.4	27.7	460	23	AW677090
36	63.4	27.7	461	23	AW286858
37	63.4	27.7	462	21	AW286839
38	63.4	27.7	463	21	AW286879
39	63.4	27.7	472	36	BE497862
40	63.4	27.7	516	35	BE363025
41	63.4	27.7	519	23	AW563648
42	63.4	27.7	530	35	BE363077
43	61.8	27.0	354	36	BE517810
44	61.8	27.0	454	36	BE495944
45	61.8	27.0	474	36	BE497880

## ALIGNMENTS

RESULT 1  
 BE022117

LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BE022117 561 bp mRNA  
 sm67f12.y1 Gm-c1028 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-c1028-9000 5' similar to TR:022111 022111 6-PHOSPHOGLUCONATE  
 DEHYDROGENASE ; mRNA sequence.  
 BE022117  
 BE022117.1 GI:8284549  
 EST.  
 soybean.

Glycine max

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE  
 AUTHORS

1 (bases 1 to 561)  
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers  
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
 R., Waterston, R. and Wilson, R.  
 Public Soybean EST Project

TITLE  
 JOURNAL  
 COMMENT

Unpublished (1999)  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com

High quality sequence stop: 398.

Location/Qualifiers  
 1..561  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1028-9000"  
 /clone\_lib="Gm-c1028"  
 /tissue\_type="roots of 'Supernod' plants"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2:  
 XhoI; The mRNA was isolated from roots of Glycine max  
 'Supernod' plants generously donated by Dr. Gary Stacey.  
 The seedlings were inoculated with Bradyrhizobium  
 japonicus, strain USDA110 prior to harvest. Stratagene's  
 cDNA synthesis kit (catalog number 200401) was used to  
 synthesize the cDNA. First-strand synthesis was performed  
 with 5-methyl dCTP, hence the ligated cDNA was  
 hemimethylated. A modification of Stratagene's  
 first-strand synthesis primer was used. An 'anchor'  
 nucleotide (V-A-C, or G) was added to the 3' end of the  
 primer (GAGAGAGAGAGAGAGAGACTAGTCGAG(T)18V) to anchor  
 the primer at the 5' end of the poly(A) tract. After  
 second-strand synthesis, the cDNA ends were filled in with  
 cloned Pfu DNA polymerase, ligated to EcoRI adapters and  
 subsequently phosphorylated. The XhoI site within the  
 first-strand synthesis primer was then restricted by  
 digestion with XhoI; all XhoI sites in the cDNA would be  
 protected by their hemimethylated status. The cDNA  
 constructs were size-fractionated with a 500bp cutoff,  
 using GibcoBRL life technologies' cDNA size fractionation  
 column. The column eluent was then ligated into  
 Stratagene's pBluescript II XR Predigested vector  
 (pBluescript II SK(+)) that has been digested with EcoRI  
 and XhoI, and phosphorylated by Stratagene. Both the  
 white and blue colonies appear to contain recombinant  
 plasmids with cDNA inserts, based on size (n=25). This  
 library was constructed by Dr. Paul Keim and Dr. Virginia  
 Coryell."

FEATURES  
 source

BASE COUNT 158 a 115 c 130 g 158 t  
 ORIGIN

Query Match	60.9%;	Score 139.4;	DB 32;	Length 561;
Best Local Similarity	81.2%;	Pred. No. 7.6e-33;		
Matches 186;	Conservative	0;	Mismatches 41;	Indels 2; Gaps 2;
1	cagacgtatttttctgctcatttgcatttccaaatttcagagattaattatcgctcaacc	60		
7	CAGACCTTAATTTTCTCTCATTCCTTCCAATTTTCAGGAATCAATTATG-GGTCAACC	65		
61	cacaacaagaataggccttgcgtgattgcgtgttaatgggcaaatctggcactcaatat	120		
66	CTCACAGAATAGGCGCTGTGTGGACTGGCTGTTATGGCGCAAAATCTAGCACATCAATAT	125		
121	tgctgaaaaggcttcccaattccggttacaacgggaaccatttccaaggtattgggc	180		
126	TGC-TGAGAAAGGCTTCCCATTTTCGTGTTATACCGGACCACTTCACAGTTGATGAGA	184		
181	cataagcgaagcaaacaccaggaagaaaccttcataattatggggaaaca	229		
185	CTGTAGACGAGCAACAGGAAGAAATCTCCAGTTTATGGGTACCA	233		

RESULT	2	
AW719787		
LOCUS	532 bp	mrna
DEFINITION	LjNEST10A4r Lotus japonicus nodule library 5 and 7 week-old Lotus japonicus cDNA 5', mRNA sequence.	EST 19-APR-2000
ACCESSION	AW719787	
VERSION	AW719787.1	GI:7614307
KEYWORDS	EST.	
SOURCE	Lotus japonicus.	
ORGANISM	Lotus japonicus	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.	
AUTHORS	1 (bases 1 to 532)	
TITLE	Colebatch,G., Freund,S., Trevasakis,B and Udvardi,M.	
JOURNAL	Lotus japonicus root nodule ESTs: tools for functional genomics	
COMMENT	Unpublished (2000)	
	Contact: Udvardi MK	
	Molecular Plant Nutrition	
	Max Planck Institute of Molecular Plant Physiology	
	Am Muehlenberg 1, 14476 Golm, Germany	
	Fax: 49 331 567 8250	
	Email: udvardi@mpimp-golm.mpg.de	
	Seq primer: T7	
	High quality sequence stop: 532.	
FEATURES	Location/Qualifiers	
	1	532
	source	

```

/organism="Lotus japonicus"
/cultivar="Gifu (B-129)"
/db_xref="taxon:34305"
/clone_lib="Lotus japonicus nodule library 5 and 7
week-old"
/dev_stage="5 and 7 week-old plants"
/note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
Site_2: NotI; The library was prepared using mRNA
extracted from nodules of 5 and 7 week-old Lotus plants.
Nodules were induced by, and contained Mesorhizobium
strain R7A."
BASE COUNT      151 a      102 c      137 g      140 t      2 others
ORIGIN

Query Match      49.5%; Score 113.4; DB 24; Length 532;
Best Local Similarity 74.2%; Pred. No. 8.5e-25;
Matches 170; Conservative 0; Mismatches 56; Indels 3; Gaps
2;

QY      3  gacctatttttcgtcattgtcttcaaatccaggagatttaattatgcgtcaaccca 62
|||||  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2  GACCTCAATTTGCTGATTTCGATTCAATTAATTCAGAAATAAATTATGGCTCAACCCAA 61
|||||  ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Qy	63	c--aacaagaaatagcctctgctggattggctgtaatggcgaaatctggcactcaatat	120
Db	62	TCATTACGAGAATAGCCCTTGCTGGACTGGCTGTTATGGCCAAAATCTTGCACCTCAATAT	121
Qy	121	tgctgaaaaagggctcccaattccggtttaacaacggcaaccatttccaaaggttattgggc	180
Db	122	TGC-TGAGAAAGGCTTCCCATTTCTGTTTACAACCGAACACATCCCAAGGTTGATGAGA	180
Qy	181	cataagacgaagcaaacacgagaaggaaccccttcaattatggggaaacaa	229
Db	181	CAGTTGACGGGCAAAACAGGAAGGAACCTTCCAGTGTATGGCTACCA	229

## RESULT 3

AV406682	430 bp	young	EST	23-MAY-2000
LOCUS	AV406682	Lotus japonicus	young plants (two-week old)	Lotus
DEFINITION	japonicus	cDNA clone MWL008a02_f 5', mRNA	sequence.	
ACCESSION	AV406682			
VERSION	AV406682.1	GI:7719536		
KEYWORDS	EST.			
SOURCE	Lotus japonicus.			
ORGANISM	Lotus japonicus			
REFERENCE	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.			
AUTHORS	Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.			
TITLE	Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus			
JOURNAL	DNA Res. 7 (2), 127-130 (2000)			
MEDLINE	20277479			
COMMENT	Contact: Yasukazu Nakamura The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/			
FEATURES	Location/Qualifiers			
SOURCE	1..430			

Query Match	48.2%	Score 110.4	DB 18	Length 430
Best Local Similarity	73.9%	Pred. No. 6.8e-24		
Matches 167	Conservative 0	Mismatches 56	Indels 3	Gaps 2
Qy	6	cttatttttctgcattgcttcaaaatttcagagattaattatgcgtcaaccca--q 63		
Db	1	CTCAATTTTGGTGTATTTCGATTCAATTTTCAGGAATAAATATGGCTCAACCAATCT 60		
Qy	64	aacagaataggcccttgctggattgggtgttaattgggcaaaattcggcactcaatttgc 123		
Db	61	TACGAGAAATAGGCCCTTGCTGGACTTGGCTGTATTGGGCCAAAATCTTCGACTCAATATTGC 120		
Qy	124	ttagaaagggtcttcccaattccggttaacaacggaaccatttccaagttattgggccat 183		
Db	121	-TGAGAAAGGCTTTCCTCATTCCTGTTTACAAACCGAACCATCCAGGTTGATGAGACAG 179		
Qy	184	aagacgaagcaaacaccaggaaggaaaccttcaattatggggaaaca 229		
Db	180	TTGAAACGGGCAAAACAGGAAGGAAACCTTCACGTGTATGGCTACCA 225		

RESULT 4

AW695821

[illegible]

QY 53 gctcaaccacaacaagaataggccttgctggattggctgttaatgggcacaaatctggca 112





```

Db 2 CTTCAATTTTCAGGAATACATTATG-GCTCAACCACTTACAAGANTAGGCTTGCGGGA 60
QY 86 ttgctgttaattgggcaaaatctgcactcaatattgttgaaaggcttcccaattcc 145
Db 61 CTGGCTGTTATGGGCAAAATCTCGCACTCAATATTGC-CGACAAAGGATTCCCAATTC 119
QY 146 ggttaacaacgaaccatttccaaggatttggccataagacgaacaaaccagaag 205
Db 120 TGTATTAAACAGAACATCAAAAGTTGACGAAACTGTGGAACGACGAAACAGGAGG 179
QY 206 aaaccttcaatt 217
Db 180 AAATCTCCACT 191

RESULT 8
LOCUS AL382847 420 bp mRNA EST 03-AUG-2000
DEFINITION MbC10C10F1 MtBC Medicago truncatula cDNA clone MtBC10C10 T3, mRNA
sequence.
ACCESSION AL382847
VERSION AL382847.1 GI:9682598
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 420)
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
,V. and Gamas,P.
MEDICAGO TRUNCATULA ESTs from endomycorrhizal roots
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
Mt-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Mtruncatula.html).
Location/Qualifiers
1. 420
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBC10C10"
/clone_lib="MtBC"
/tissue_type="arbuscular mycorrhiza"
/dev_stage="harvested 3 weeks post inoculation with Glomus
intraradices"
/notes="vector: pBluescript PSK; Site.1: EcoRI; Site.2:
XhoI; M. truncatula sterilised seeds were germinated for
72h at 25 C, before transplanting into a 1/3 Epouisses soil
: 2/3 calcined Terragreen mix in the presence of onion
root fragments colonized by the arbuscular mycorrhizal
fungus Glomus intraradices (Schenck & Smith, isolate LPA8
). The plants were watered every day and twice a week with
a modified nutrient Long Ashton solution without phosphate
but with a high level of nitrate. After 3 weeks RNA was
extracted from whole root systems. cDNA was prepared from
polyA+ enriched RNA. The cDNA was directionally ligated
into Uni-zap XR vector from Stratagene and packaged using
Gigapack Gold packaging extracts. Plasmids containing cDNA
inserts were mass-excised from phage stocks using ExAssit
helper phage and propagated in SOLR cells. Clone ordering
and sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France). Note : EST may be of
fungal origin."
BASE COUNT 128 a 83 c 102 g 106 t 1 others
ORIGIN

```

```

Query Match 39.1%; Score 89.6; DB 14; Length 420;
Best Local Similarity 74.6%; Pred. No. 1.9e-17;
Matches 138; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

QY 33 tticaggagataattatcgctcaaccacacaaagaatagccttgcgtgattggctg 92
Db 1 TTTCAAGGAATACATTATG-GCTCAACCACTTACAAGANTAGGCTTGCGGACTGGCTG 59
QY 93 ttaatgggcaaaatctggcactcaatattgttgaaagggttcccaattccggttaac 152
Db 60 TTATGGGCCAAAATCTCGCACTCAATATTGC-CGACNAAGGATTCCCAATTTCTGTTAT 118
QY 153 aacggaaaccatttccaaggatttggccataagacgaacaaaccagaagaaacctt 212
Db 119 AACAGAAACACATCAAAAGTTGACGAAACTGTGGAACGACGAAAAACAGAGGGAATCTT 178
QY 213 caatt 217
Db 179 CCACT 183

RESULT 9
LOCUS AL367204 337 bp mRNA EST 03-AUG-2000
DEFINITION MtBA12H03F1 McBA Medicago truncatula cDNA clone MtBA12H03 T3, mRNA
sequence.
ACCESSION AL367204
VERSION AL367204.1 GI:9666957
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 337)
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
,V. and Gamas,P.
MEDICAGO TRUNCATULA ESTs from nitrogen-starved roots
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
Mt-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Mtruncatula.html).
Location/Qualifiers
1. 337
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBA12H03"
/clone_lib="MtBA"
/tissue_type="root tips"
/dev_stage="harvested after 3 days of N-starvation"
/notes="Vector: pBluescript PSK; Site.1: EcoRI; Site.2:
XhoI; Plants were grown in an aeroponic chamber for 14
days on nitrogen-rich medium followed by 3 days on N-free
medium. RNA was extracted from root tips (1-3 cm). cDNA
was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zapXR vector from
Stratagene and packaged using Gigapack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using ExAssit helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."
BASE COUNT 108 a 71 c 72 g 86 t
ORIGIN

```





```

JOURNAL DNA Res. 7 (2), 127-130 (2000)
MEDLINE 20277479
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
    /organism="Lotus japonicus"
    /db_xref="taxon:34305"
    /clone="WM200c07_r"
    /clone_lib="Lotus japonicus young plants (two-week old)"
    /dev_stage="young plants (two-week old)"
    /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
    XhoI; isolate=Miyakojima MG-20"
BASE COUNT 114 a 80 c 89 g 96 t
ORIGIN

Query Match 37.4%; Score 85.6; DB 18; Length 379;
Best Local Similarity 71.6%; Pred. No. 3.2e-16;
Matches 126; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 54 ctcaacccacaagaatagccttgcctgattggtggttaattggtgcaaaatctggcac 113
Db 14 CTCCAACCCCTACAAGATTGGCCTTGTGGATTGGCTGTGATGGGCCCAAAACCTTGCTC 73

QY 114 tcaatatgtctgaaagggtctcccaattccggttaacaacggaaccatttccaaggtt 173
Db 74 TCAACATTGC-TGATAAGGGCTTCCCATTTCTGTTTATAACCGAACCATCAAGGTT 132

QY 174 attgggccataagcagcaaccaggaagaaacctcaatttatggggaacaa 229
Db 133 GATGAGACTGTCAGACGACGAAAAAAGAGAAATCTTCTGTGTATGGCTACCA 188

RESULT 15
AV425788 394 bp mRNA EST 23-MAY-2000
LOCUS AV425788 Lotus japonicus young plants (two-week old) Lotus
DEFINITION japonicus cDNA clone MM059a03_r 5', mRNA sequence.
ACCESSION AV425788
VERSION AV425788.1 GI:7784076
KEYWORDS EST.
SOURCE Lotus japonicus.
ORGANISM Lotus japonicus
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE Generation of 7137 non-redundant expressed sequence tags from a
legume, Lotus japonicus
JOURNAL DNA Res. 7 (2), 127-130 (2000)
MEDLINE 20277479
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
    /organism="Lotus japonicus"
    /db_xref="taxon:34305"
    /clone="WM059a03_r"
    /clone_lib="Lotus japonicus young plants (two-week old)"
    /dev_stage="young plants (two-week old)"
    /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
    XhoI; isolate=Miyakojima MG-20"
BASE COUNT 115 a 85 c 93 g 101 t
ORIGIN

```

```

Query Match 37.4%; Score 85.6; DB 18; Length 394;
Best Local Similarity 71.6%; Pred. No. 3.2e-16;
Matches 126; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

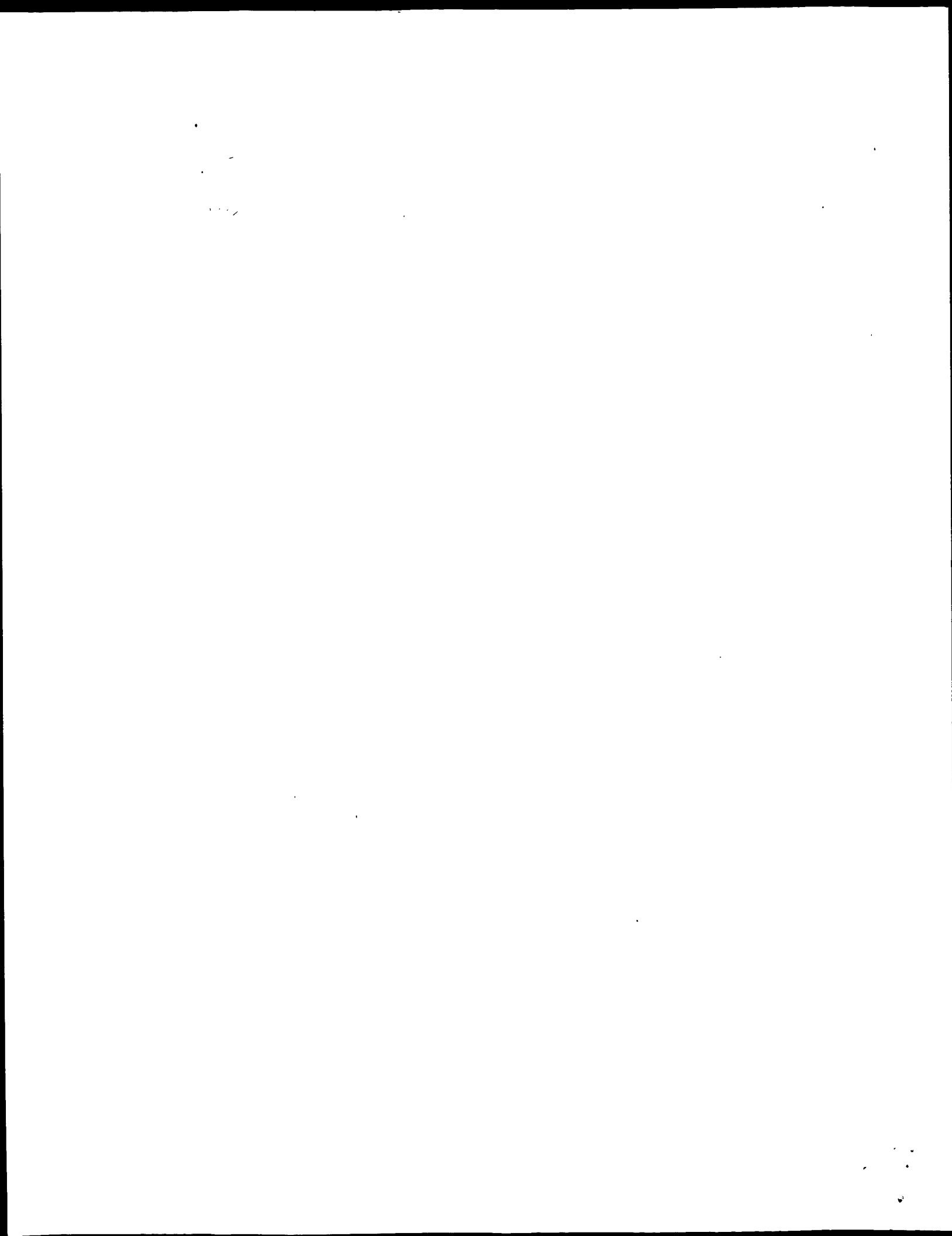
QY 54 ctcaacccacaagaatagccttgcctgattggtggttaattggtgcaaaatctggcac 113
Db 11 CTCCAACCCCTACAAGATTGGCCTTGTGGATTGGCTGTGATGGGCCCAAAACCTTGCTC 70

QY 114 tcaatatgtctgaaagggtctcccaattccggttaacaacggaaccatttccaaggtt 173
Db 71 TCAACATTGC-TGATAAGGGCTTCCCATTTCTGTTTATAACCGAACCATCAAGGTT 129

QY 174 attgggccataagcagcaaccaggaagaaacctcaatttatggggaacaa 229
Db 130 GATGAGACTGTCGAAACGACGAAAAAAGAGAAATCTTCTGTGTATGGCTACCA 185

```

Search completed: November 4, 2000, 11:52:06  
 Job time: 18243 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:30:03 ; Search time 4075.18 Seconds  
(without alignments)  
281.871 Million cell updates/sec

Title: US-09-300-482-4  
Perfect score: 263  
Sequence: 1 gaagcacttttgatgtgc.....agnngnacntnnnnnganna 263

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_bal:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pl1:\*
- 8: gb\_pl2:\*
- 9: gb\_pr1:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: em\_fun:\*
- 13: em\_hum1:\*
- 14: em\_hum2:\*
- 15: em\_in:\*
- 16: em\_om:\*
- 17: em\_or:\*
- 18: em\_ov:\*
- 19: em\_pat:\*
- 20: em\_ph:\*
- 21: em\_pl1:\*
- 22: em\_ro:\*
- 23: em\_sts:\*
- 24: em\_sy:\*
- 25: em\_un:\*
- 26: em\_vi:\*
- 27: gb\_htg1:\*
- 28: gb\_htg2:\*
- 29: gb\_in1:\*
- 30: gb\_in2:\*
- 31: em\_bal:\*
- 32: em\_ba2:\*
- 33: em\_hum3:\*
- 34: em\_hum4:\*
- 35: gb\_pr4:\*
- 36: gb\_htg3:\*
- 37: gb\_htg4:\*
- 38: gb\_htg5:\*
- 39: gb\_hcg6:\*
- 40: gb\_hcg7:\*
- 41: em\_htg1:\*
- 42: em\_htg2:\*
- 43: em\_htg3:\*

- 44: em\_hum5:\*
- 45: gb\_pl3:\*
- 46: gb\_pr5:\*
- 47: gb\_htg8:\*
- 48: gb\_htg9:\*
- 49: gb\_htg10:\*
- 50: gb\_htg11:\*
- 51: gb\_htg12:\*
- 52: gb\_htg13:\*
- 53: gb\_htg14:\*
- 54: gb\_in3:\*
- 55: gb\_htg15:\*
- 56: gb\_htg16:\*
- 57: gb\_htg17:\*
- 58: em\_htg4:\*
- 59: em\_htg5:\*
- 60: em\_htg6:\*
- 61: em\_hcg7:\*
- 62: em\_hum6:\*
- 63: gb\_htg18:\*
- 64: gb\_htg19:\*
- 65: gb\_ba3:\*
- 66: em\_htg8:\*
- 67: em\_htg9:\*
- 68: em\_htg10:\*
- 69: em\_htg11:\*
- 70: em\_htg12:\*
- 71: em\_htg13:\*
- 72: em\_htg14:\*
- 73: em\_htg15:\*
- 74: em\_htg16:\*
- 75: em\_htg17:\*
- 76: em\_htg18:\*
- 77: em\_htg19:\*
- 78: em\_htg20:\*
- 79: em\_htg21:\*
- 80: em\_htg22:\*
- 81: em\_htg23:\*
- 82: gb\_pr6:\*
- 83: gb\_pr7:\*
- 84: gb\_htg20:\*
- 85: gb\_htg21:\*
- 86: gb\_htg22:\*
- 87: gb\_htg23:\*
- 88: gb\_ro:\*
- 89: gb\_sts1:\*
- 90: gb\_sts2:\*
- 91: gb\_sy:\*
- 92: gb\_un:\*
- 93: gb\_vil:\*
- 94: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
C	1	76.2	29.0	95865	7	AC003970	AC003970 Arabidops
	2	68.6	26.1	2058	45	STG6PDHPI	X83923 S.tuberosum
	3	66.6	25.3	1396	45	SO000184	AJ000184 Spinacia
	4	66.6	25.3	2049	45	SO000182	AJ000182 Spinacia
	5	62.2	23.7	1869	45	NTTPG18	AJ001772 Nicotiana
C	6	61.4	23.3	10195	65	U32737	U32737 Haemophilus
	7	61	23.2	1630	8	ATG6PDHES	X84229 A.thaliana
	8	60.6	23.0	1942	45	NTTPG16	AJ001771 Nicotiana
	9	60	22.8	1975	8	ATJ00135	AJ001359 Arabidops
	10	59	22.4	1772	8	ATG6PDH15	X84230 A.thaliana
C	11	58.4	22.2	220373	64	AL158831	AL158831 Homo sapi
	12	56.8	21.6	1976	8	CCA6246	AJ006246 Cyanidium





5504. .7028  
/gene="F14J9.3"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AAC33197.1"  
/db\_xref="GI:3482912"  
/translation="MHHLSIVLAILASSVAEAFLEDEKTOORVNGICKOTMDTRFCS  
SLILKNLTPFASKEINLNVSVSAEPAANTYFFISTLLNAGDERPDQACAEAVA  
LYNSATKAVFFQKQYVSKIVNIEKKVSMVADICKTDENVLGVIQINPLEKKNQTKI  
LLSMQIVSHVSVS"  
complement(7954. .10925)  
/gene="F14J9.4"  
complement(join(7954. .8155,8251. .8402,8490. .8648,  
8762. .9014,9101. .9217,9294. .9356,9577. .9606,9895. .9938,  
10404. .10444,10829. .10925))  
/note="F14J9.4"  
/note="Similar to MtN21, gi|2598575, Megicaco truncatula  
nodulation induced gene"  
/codon\_start=1  
/protein\_id="AAC33198.1"  
/db\_xref="GI:3482913"  
/translation="MAKSDMLPFLAMVLVVOIGVAGMNTSKMAEAEDEIIFINIV  
KMYKNVKKKTKWCRAIFICQDPTKPKTKTLRLVQVFCSTGATGNVLIFV  
GLNSPTIACALNLLPAPVFLAALFRGETVGIKASQKAVIGTLVVICGAMVLS  
FYHGHTIGESKHWAYANNTKHGSSGNFPLGFLPMAAAVSWAWFIIOTKM  
SETAAPTSTLLMCLMGISQCGAIALISDHTISDMSLPLRFSIALYAGVVASALA  
FCLSMAQRKPLVSVFSPFLVLLVVAIFSWALLEKLVTFMGSAVLVIGLYCVL  
WGDREYSEKEEREKVKQONHKVKSBNESNEDIESRLPVASSGNGSTRSP"  
12867. .15018  
/gene="F14J9.5"  
join(12867. .13122,13720. .13911,13990. .14156,14269. .14533,  
14786. .15018)  
/note="F14J9.5"  
/note="Similar to nodulins and lipase; location of EST  
E6C2T7, gb|AA042309, similar to nodulins gi|3328240,  
gi|2129854 and others and lipase, gi|2129636"  
/codon\_start=1  
/protein\_id="AAC33199.1"  
/db\_xref="GI:3482914"  
/translation="MATLSHSHSFLVLLPFLILQNLAVAGGQVPPVIFNFGDS  
NSDTGVLGYSIGLNGSRFSFORSTGLSDGLRVIDFLQSLNTLLNLYLDSLV  
GSKFQNGANVIGSSTLPYVPFALNOLMOPHPSKALEIASISDLPKEMMIGES  
GFNVALMIDIGONDIADESGLYSRVVYLPVWISKAIKILYDEGGRKEVWH  
NTGLCLQKLSMWSHSGEDKHGCLVATNAAKLFNEGLDHMCRLRTELKEANIV  
VDYIAKYDLIANSNNGYFEKPLMACCGYGGPPYNNVITCGNGSKSCDEGSRFIS  
WDGIHTETANAIAMKVLMSQHSTPTTFFHFCGG"  
complement(15279. .16596)  
/gene="F14J9.6"  
complement(join(15279. .15796,15879. .16037,16122. .16252,  
16328. .16468,16571. .16596))  
/note="F14J9.6"  
/note="Similar to 12-oxophytodienoate reductase,  
gi|2765083 and old-yellow-enzyme homolog, gi|2232254"  
/codon\_start=1  
/protein\_id="AAC33200.1"  
/db\_xref="GI:3482915"  
/translation="MKNELTRIYVAPWARMRSNTIQQPHVLYCQRTTPGGLLI  
SPATGSETAMAYQNPGLWKEQIEANKPIVDVSHGGIFFCOLWHRAGVHQDCQ  
PNGESPVSTDKPAPDSNEFTPRRLRDEIPTIINDFLAARNATEAGFDGVEIHH  
GAGHYLDIFMDKSDVNDRTSYGSLNRCFALQVIEAVSKEIGPDVRVGLRSPFAD  
YMGSDTDRKRLGLMAKSLNRFILYCHMIIEPMKTVSEIFECESEITPMRNFNCT  
FVAVGTYREDGNKVAEGRTDLVAGRLFANPDLPRFELNAPLNK"  
17046. .19163  
/gene="F14J9.7"  
17046. .19163  
/note="F14J9.7"  
/note="Hypothetical protein; Similar to Arabidopsis  
selenium-binding protein, gi|2244760 and Arabidopsis  
hypothetical proteins gi|3033399, gi|2464864, gi|2244839  
and several others"  
/codon\_start=1  
/protein\_id="AAC33201.1"  
/db\_xref="GI:3482916"





Query Match	23.7%	Score 62.2	DB 45	Length 1869	
Best Local Similarity	51.2%	Pred. No. 4.2e-06			
Matches 127	Conservative	0	Mismatches 121	Indels	Gaps 0
QY	6	acttttggttggtgcattgtcttcttgcaagcagtgctcagaccacgaagggatggaatcg	55		
DB	681	ATTCTGGTGGATGGTGGAGATGTCGAAGTCTTAAAGCTTTCACAAAGTGGGTGGCAAG	740		
QY	66	cataatattgagaagccattggcttcttgatgcacttcttcccataggtgacacaata	125		
DB	741	AGTCATGTGTGAGAAACCGTTTGGTCGTGACCTAGATCATCTAGTAGACTGACCAGATG	800		
QY	126	tcttcttcaaaacttccagaaaagcaaatatagaaatgancatctactagggaagaa	185		
DB	801	CCTAAAGAAAGTATCTAACTGAGGAGCAAAATATTCGGAATTGACCAFTACTTGGGAAAGGA	860		
QY	186	tcncaadnaaatcctncaggtttaagggttcacaaannnagnttttagccacccttngag	245		
DB	861	ACTGTGTGAGATCTCTCAGTCTTCTGTTCTCGAATCTTGCTTTGAGCCCTCTGTGTC	920		
QY	246	nngnaent	253		
DB	921	CAGAAACT	928		
RESULT	6				
U32737/c					
LOCUS	U32737	10195 bp	DNA	BCT	29-MAY-1998
DEFINITION	Haemophilus influenzae Rd section 52 of 163 of the complete genome.				
ACCESSION	U32737	L42023			
VERSION	U32737.1	GI:1573536			
KEYWORDS	Haemophilus influenzae Rd.				
SOURCE	Haemophilus influenzae Rd				
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.				
REFERENCE	1 (bases 1 to 10195)				
AUTHORS	Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kervilave,A.R., Bult,C.J., Tomb,J., Dougherty,B.A., Merrick,J.M., McKenney,K., Sutton,G.G., Fitzhugh,W., Fields,C.A., Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A., Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E., Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M., Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L., Goughagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M., Smith,H.O. and Venter,J.C.				
TITLE	Whole-genome random sequencing and assembly of Haemophilus influenzae Rd				
JOURNAL	Science 269 (5223), 496-512 (1995)				
MEDLINE	95350630				
REFERENCE	2 (bases 1 to 10195)				
AUTHORS	Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V.				
TITLE	Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli				
JOURNAL	Curr. Biol. 6 (3), 279-291 (1996)				
MEDLINE	96398784				
REFERENCE	3 (bases 1 to 10195)				
AUTHORS	White,O., Clayton,R.A., Kervilave,A.R. and Fleischmann,R.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
REFERENCE	4 (bases 1 to 10195)				
AUTHORS	White,O., Clayton,R.A., Kervilave,A.R. and Fleischmann,R.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
REFERENCE	The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes fields of the corresponding H. influenzae genes				
REMARK					

PID:145690 percent identity: 36.07; identified by sequence similarity; putative	
/codon_start=1	
/transl_table=11	
/product="sulfitase synthesis pathway protein (cysQ)"	
/protein_id="AAC22214.1"	
/db_xref="GI:1573544"	
/translation="MTEADLFVSQFTEKLTALFPNPVLSENCHISFEERNKWEY	
Query Match	
Best Local Similarity 52.1%; Score 61.4; DB 65; Length 10195; *	
Matches 110; Conservative 0; Mismatches 101; Indels 0; Gaps 0;	
QY	53 agggatggaatccataatttgagaagcattggctttgatgcactttctccata 112
Db	
QY	6327 ATGGCTGGAACGATTATTCGTTGAAACACCTTTGGTTACGATGAAACACGCGCAA 6268
Db	
QY	113 ggcagacacaatatcttcttcaaacatttcaggaaagcaaatatagaaattgancatc 172
Db	
QY	6267 CATTAGACGTFACAAATCCACCGTTTCTTTGAGAACACCAAAATTTATCGTATCGATCAT 6208
QY	173 tactagaaggaatcncagtnaaatctcncaggtttaagggttccaaannagntttg 232
Db	
QY	6207 ATTAGGTAAAGAAACCGTTTCAAACTTGCTCGTATTACGTTTTCAAATGGTTGGTTG 6148
Db	
QY	233 agccactngagngnagcnnnnnganna 263
Db	
QY	6147 AACCGCTTTGGAACCGTAAATTCATTGATTA 6117
Db	
RESULT 7	
ATG6PDHE5 1630 bp mRNA PLN 31-JAN-1996	
LOCUS A.thaliana mRNA for glucose-6-phosphate dehydrogenase (clone E5).	
DEFINITION X84229	
ACCESSION X84229.1 GI:1166404	
VERSION 96phd gene; glucose-6-phosphate dehydrogenase.	
KEYWORDS thale cress.	
SOURCE Arabidopsis thaliana	
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsis.	
1 (bases 1 to 1630)	
Fink,A., Greppin,H. and Tacchini,P.	
Nucleotide sequence of a cDNA encoding the glucose-6-phosphate dehydrogenase from Arabidopsis thaliana	
Unpublished	
2 (bases 1 to 1630)	
Fink,A.F.A.	
Direct Submission	
Submitted (23-JAN-1995) A.F.A. Fink, Lab de Biochimie et Physiologie, Vitales, Universiti de Geneve, 3 place de l'Universiti, 1211 Geneva, SWITZERLAND	
Location/Qualifiers	
1. 1630	
/organism="Arabidopsis thaliana"	
/cultivar="Columbia"	
/db_xref="taxon:3702"	
/dev_stage="adult"	
/tissue_type="leaves, stems, some flowers and roots"	
/cell_line="A.thalian Co"	
/clone_lib="lambda ZAP"	
/clone="E5"	
<1. 1630	
/gene="G6PDH"	
1. 1630	
/gene="G6PDH"	
<1. 1480	
/gene="G6PDH"	
/EC_number="1.1.1.49"	
/codon_start=2	
/product="glucose-6-phosphate 1-dehydrogenase"	
/protein_id="CAA59011.1"	
mRNA	
gene	
CDS	

FEATURES	Location/Qualifiers
source	1. .1942
	/organism="Nicotiana tabacum"
	/cultivar="Samsun NN"
	/db_xref="taxon:4097"
	/cell_type="mesophyll"
	/tissue_type="young sink leaves"
sig_peptide	41..232
mRNA	41..1801
	/note="TPG16"
	/product="glucose-6-phosphate dehydrogenase"

mat_peptide	233..1798	/product="glucose-6-phosphate dehydrogenase"
BASE_COUNT	584 a	353 c 442 g 563 t
ORIGIN		

Query Match	23.0%	Score	60.6	DB	45	Length	1942
Best Local Similarity	50.8%	Pred. No.	1.1e-05				
Matches	126	Conservative	0	Mismatches	122	Indels	0

[illegible]

RESULT	9	17-AUG	PLN	mRNA	PLN	mRNA
ATJ00135						
LOCUS	ATJ00135	1975 bp				

DEFINITION	Arabidopsis thaliana cDNA encoding plastidic glucose 6-phosphate dehydrogenase.
ACCESSION	AJ001359
VERSION	AJ001359.1 GI:3021304
KEYWORDS	glucose-6-phosphate dehydrogenase.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta

REFERENCE AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1975) Wendt, U.K., Hauschild, R., Lange, C., Pietersma, M., Wendersma, A. von Schaewen, A.	Evidence for functional convergence of redox regulation	
euphyllipolites; spermatophyta; magnoliophyta; eudicotyledons; rosidae; eurosids II; brassicales; brassicaceae Arabidopsis.	isoforms of cyanobacteria and higher plants	Plant Mol. Biol. 40 (3), 487-494 (1999)

MEDLINE	99364543
REFERENCE	2 (bases 1 to 1975)
AUTHORS	von Schaewen, A.
TITLE	Direct Submission
JOURNAL	Submitted (18-SEP-1997) von Schaewen A., Plant Physiology University Osnabrueck, Barbarastr. 11, D-49069 Osnabrueck
FEATURES	location/Qualifiers
source	1..1975 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /tissue_type="Above ground parts"

## CDS

```

71. 1801
/codon_start=1
/product="plastidic glucose-6-phosphate dehydrogenase"
/protein_id="CAA04696.1"
/db_xref="GI:3021305"
/db_xref="SPTREMBL:O65577"
/translation="MATHSMIIPSPSSSSSLATAAPFKETLPFRSLTPFRKSLF
SOVLRPFAEKHSOLDTSNCATNFASQDSGDLTEEHVTKGESTLSITVVGASGDL
AKKIFALFALFEGCLPDQFVSFYARTKLTHEELRDMISSTLCRIDQREKCGDK
MEQCLKCFYHSGYNSSEDFAEALNKKLKEKAGKISNRLYILSPNLFVDVRCASLRAS
LRASNGWTRVIVKEKPGFGRDSESSGELTRCLQYLTEQIFRIDHYLKGELVENLSV
LRFNLVPEPLWSRNYIRNVQLIFSEDFEGTGRGGYFDQYGIIRDIMQHLLQILALF
AMETPSVLDADIRSEKVKVLRSMKPLRLDEDVVGQYKGNKGGKTYPTDDPTVFN
HSLPTTAAAMEINNAWDGVPFLMKAGKALHTRGAEIRVQFRHVPGLNLYKKSFTAT
LDNATNELVIRVQDEGIYLRINNKVPGLMRLDRDLNLLYRSRYPREIPDAYERLL
LDLAEARRLFIKSDLDAAWDLFTALKELEEKIIPELYPGSRGPVGAHYLASKY
NVRWGDLGEA"

```

BASE COUNT 583 a 382 c 447 g 563 t

Query Match 22.8%; Score 60; DB 8; Length 1975;  
 Best Local Similarity 52.5%; Pred. No. 1.5e-05;  
 Matches 114; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

QY 37 agtgcagaccaggaagggaatcgcaataatttgcagaccatttgcttgat 96
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 AGAGCCTCCTCAGAGATGCGTGACACAGGTCATTGTGAAAGCCATTGGCCGTGAC 787

QY 97 gcaatttctccatagctgacacatatcttcttcaaaccttcaggaaagcaata 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 TCTGAATCATCCGGAGAGTTAACTAGATGCTGAAACAGTATCTTACAGAGGAGCAATC 847

QY 157 tatgaattgancatctactaggaagaatcncagtnaaatcctcaggttaagggtt 216
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 TTCAGGATTGATCATCTATTGGGAAAGGAGCTGTGTGAGAACCTTTCAGTCTCCGATTC 907

QY 217 tcaaanngnntttgagccaccttngagngnnaact 253
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 908 TCAAACTCTGTTTCGAGCCTCTGTGTCAAGAAAT 944

```

RESULT 10  
 ATG6PDH15 1772 bp mRNA PLN 15-AUG-1997  
 LOCUS A.thaliana mRNA for glucose-6-phosphate dehydrogenase (clone EL5).  
 DEFINITION X84230

VERSION 84230.1 GI:1174335  
 KEYWORDS g6pdh gene; glucose-6-phosphate dehydrogenase.  
 SOURCE thale cress; Arabidopsis thaliana

ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 Rosidae; Capparales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 1772)  
 Fink, A.F.A.

Direct Submission  
 Submitted (23-JAN-1995) A.F.A. Fink, Lab de Biochimie et de  
 Physiologie, Vitis, Universit  de Geneve, 3 Place de  
 l'Universit , 1211 Geneva, SWITZERLAND  
 2 (bases 1 to 1772)  
 Fink, A., Greppin, H. and Tacchini, P.

Nucleotide sequence of a cDNA encoding the glucose-6-phosphate  
 dehydrogenase (Genbank X84230) from Arabidopsis thaliana  
 Plant Physiol. 108, 1343-1343 (1995)  
 Ref [3]: Plant Gene Register PGR95-021 (1995).  
 Location/Qualifiers  
 1. 1772

ORGANISM="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /haplotype="diploid"  
 /dev\_stage="adult"

```

/tissue_type="leaves, stems, flowers and roots"
/cell_line="A.thaliana Co"
/clone_lib="lambda ZAP"
/clone="E15"
141. 1685
/gene="G6PDH"
141. 1685
/gene="G6PDH"
141. 1685
/gene="G6PDH"
/EC_number="1.1.1.49"
/codon_start=1
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="CAA59012.1"
/db_xref="GI:1174336"
/db_xref="SWISS-PROT:Q43727"
/translation="MGVATNFASQDSGDLTEEHVTKGESTLSITVVGASGLAKKK
IFPDLFALFEGCLPDQFVSFYARTKLTHEELRDMISSTLCRIDQREKCGDKMEQF
LKRCFYHSGYNSSEDFAEALNKKLKEKAGKISNRLYILSPNLFVDVRCASLRAS
SENGWTRVIVKEKPGFGRDSESSGELTRCLQYLTEQIFRIDHYLKGELVENLSVURFS
NLVPEPLWSRNYIRNVQLIFSEDFEGTGRGGYFDQYGIIRDIMQHLLQILALFAMET
PSVLDADIRSEKVKVLRSMKPLRLDEDVVGQYKGNKGGKTYPTDDPTVFNHSLT
PTAAAMEINNAWDGVPFLMKAGKALHTRGAEIRVQFRHVPGLNLYKKSFTATNLDAI
TNELVIRVQDEGIYLRINNKVPGLMRLDRDLNLLYRSRYPREIPDAYERLLDAI
EGEERLFIKSDLDAAWDLFTALKELEEKIIPELYPGSRGPVGAHYLASKYNRVW
GDLGEA"

```

polyA\_signal 1756. 1761  
 polyA\_signal 1772  
 BASE COUNT 526 a 331 c 419 g 496 t  
 ORIGIN

Query Match 22.4%; Score 59; DB 8; Length 1772;  
 Best Local Similarity 50.4%; Pred. No. 2.8e-05;  
 Matches 125; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```

QY 6 acttttgatgtgcgcattgttcagagcagtgctcagaccaggaaggatggaatcg 65
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 ATTCGTTGATGTGGTTAGGTGTCAGAGCTCTTACAGAGCTCTCAGAGAGTGGTGGACAAG 640

QY 66 cataatatttgcagaccatttgcttgatgcactttctccataggtgacacaata 125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641 GGTCATTGTGGAAAGCCATTGGTCTGCTGACTGCTGATCATCCGAGAGTTAACTAGATG 700

QY 126 tcttcttcaaaccttcaggaaagcaataatatagaaattgancatctactaggaagaa 185
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 TCTGAACACAGTATCTTACAGAGGAGCAATCTTCAGGATTGATCACTATTTCGGGAAAGCA 760

QY 186 tencagtnaaatcctcaggtttaaagggtttcaaanngnnttttgcacaccttngag 245
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 761 GCTTGTGAGACCTTTCAGTCTCGGATTCGAGATTCTCAAACTCTTGTTCGAGCCTCTGTGTC 820

QY 246 nngnaact 253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 821 AAGAAAT 828

```

RESULT 11  
 AL158831/C 220373 bp DNA HTG 22-MAY-2000  
 LOCUS Homo sapiens chromosome 9 clone RP11-564A4, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, 33 unordered pieces.  
 ACCESSION AL158831  
 VERSION AL158831.5 GI:9212620  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 220373)  
 Plumb, B.  
 Direct Submission  
 Submitted (21-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Sat Nov 4 18:11:37 2000

COMMENT

On Jul 15, 2000 this sequence version replaced gi:7939701.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: BA564A4  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 200906 bases at least Q40  
 Consensus quality: 208868 bases at least Q30  
 Consensus quality: 213166 bases at least Q20  
 Insert size: 217173; sum-of-contigs  
 Insert size: 157787; agarose-fp  
 Quality coverage: 3.4x in Q20 bases; sum-of-contigs Quality  
 coverage: 4.71x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 33 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 19404: contig of 19404 bp in length  
 \* 19405 19504: gap of 100 bp  
 \* 19505 20573: contig of 1069 bp in length  
 \* 20574 20673: gap of 100 bp  
 \* 20674 22382: contig of 1709 bp in length  
 \* 22383 22482: gap of 100 bp  
 \* 22483 30989: contig of 8507 bp in length  
 \* 30990 31089: gap of 100 bp  
 \* 31090 34726: contig of 3637 bp in length  
 \* 34727 34826: gap of 100 bp  
 \* 34827 36278: contig of 1452 bp in length  
 \* 36279 36378: gap of 100 bp  
 \* 36379 57138: contig of 20760 bp in length  
 \* 57139 57238: gap of 100 bp  
 \* 57239 61395: contig of 4157 bp in length  
 \* 61396 61495: gap of 100 bp  
 \* 61496 63183: contig of 1888 bp in length  
 \* 63184 63283: gap of 100 bp  
 \* 63284 64346: contig of 1063 bp in length  
 \* 64347 64446: gap of 100 bp  
 \* 64447 66218: contig of 1772 bp in length  
 \* 66219 66318: gap of 100 bp  
 \* 66319 67584: contig of 1266 bp in length  
 \* 67585 67684: gap of 100 bp  
 \* 67685 68875: contig of 1191 bp in length  
 \* 68876 68975: gap of 100 bp  
 \* 68976 71548: contig of 2573 bp in length  
 \* 71549 71648: gap of 100 bp  
 \* 71649 73089: contig of 1441 bp in length  
 \* 73090 73189: gap of 100 bp  
 \* 73190 81026: contig of 7837 bp in length  
 \* 81027 81126: gap of 100 bp  
 \* 81127 82976: contig of 1850 bp in length  
 \* 82977 83076: gap of 100 bp  
 \* 83077 88550: contig of 5474 bp in length  
 \* 88551 88650: gap of 100 bp  
 \* 88651 89972: contig of 1322 bp in length  
 \* 89973 90072: gap of 100 bp  
 \* 90073 94252: contig of 4180 bp in length  
 \* 94253 94352: gap of 100 bp  
 \* 94353 97066: contig of 2714 bp in length  
 \* 97067 97166: gap of 100 bp  
 \* 97167 100212: contig of 3046 bp in length  
 \* 100213 100312: gap of 100 bp  
 \* 100313 103147: contig of 2835 bp in length

\* 103148 103247: gap of 100 bp  
 \* 103248 104757: contig of 1510 bp in length  
 \* 104758 104857: gap of 100 bp  
 \* 104858 106680: contig of 1823 bp in length  
 \* 106681 106780: gap of 100 bp  
 \* 106781 108222: contig of 1442 bp in length  
 \* 108223 108322: gap of 100 bp  
 \* 108323 109526: contig of 1204 bp in length  
 \* 109527 109626: gap of 100 bp  
 \* 109627 156013: contig of 46387 bp in length  
 \* 156014 156113: gap of 100 bp  
 \* 156114 162589: contig of 6476 bp in length  
 \* 162590 162689: gap of 100 bp  
 \* 162690 189785: contig of 27096 bp in length  
 \* 189786 189885: gap of 100 bp  
 \* 189886 198960: contig of 9075 bp in length  
 \* 198961 199060: gap of 100 bp  
 \* 199061 211582: contig of 12522 bp in length  
 \* 211583 211682: gap of 100 bp  
 \* 211683 220373: contig of 8691 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1..220373  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-564A4"  
 /clone\_lib="RPCI-11.2"  
 1..19404  
 /note="assembly\_fragment:01468"  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left  
 19505..20573  
 /note="assembly\_fragment:00210"  
 fragment\_chain:1  
 20674..22382  
 /note="assembly\_fragment:00702"  
 fragment\_chain:2  
 22483..30989  
 /note="assembly\_fragment:01191"  
 fragment\_chain:2  
 31090..34726  
 /note="assembly\_fragment:00723"  
 fragment\_chain:2  
 34827..36278  
 /note="assembly\_fragment:00747"  
 fragment\_chain:3  
 36379..57138  
 /note="assembly\_fragment:00108"  
 fragment\_chain:3  
 57239..61395  
 /note="assembly\_fragment:01455"  
 fragment\_chain:3  
 61496..63183  
 /note="assembly\_fragment:01321"  
 fragment\_chain:4  
 63284..64346  
 /note="assembly\_fragment:01175"  
 fragment\_chain:4  
 64447..66218  
 /note="assembly\_fragment:00029"  
 66319..67584  
 /note="assembly\_fragment:00187"  
 /note="assembly\_fragment:00187"  
 67685..68875  
 /note="assembly\_fragment:00310"  
 68976..71548  
 /note="assembly\_fragment:00352"  
 71649..73089  
 /note="assembly\_fragment:00419"  
 73190..81026  
 /note="assembly\_fragment:00466"  
 81127..82976  
 /note="assembly\_fragment:00799"



misc\_feature 83077..88550  
 /note="assembly\_fragment:00825"  
 misc\_feature 88551..89972  
 /note="assembly\_fragment:00990"  
 misc\_feature 90073..94252  
 /note="assembly\_fragment:01020"  
 misc\_feature 94353..97066  
 /note="assembly\_fragment:01043"  
 misc\_feature 97167..100212  
 /note="assembly\_fragment:01173"  
 misc\_feature 100313..103147  
 /note="assembly\_fragment:01212"  
 misc\_feature 103248..104757  
 /note="assembly\_fragment:01234"  
 misc\_feature 104858..106680  
 /note="assembly\_fragment:01312"  
 misc\_feature 106781..108222  
 /note="assembly\_fragment:01421"  
 misc\_feature 108323..109526  
 /note="assembly\_fragment:01462"  
 misc\_feature 109627..156013  
 /note="assembly\_fragment:00918"  
 misc\_feature 156114..162589  
 /note="assembly\_fragment:00811"  
 misc\_feature 162690..189785  
 /note="assembly\_fragment:00625"  
 misc\_feature 189886..198960  
 /note="assembly\_fragment:00180"  
 misc\_feature 199061..211582  
 /note="assembly\_fragment:00696"  
 misc\_feature 211683..220373  
 /note="assembly\_fragment:01298"  
 fragment\_chain:5

Query Match 22.2%; Score 58.4; DB 64; Length 220373;  
 Best Local Similarity 52.6%; Pred. No. 4e-05;  
 Matches 110; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
 QY 45 gaccagaaggatggaatgcataatatttgaagaccatttggcttggatgcacttc 104  
 DB 91224 GAACGTAAAGGGTTTAAACCAATATATAGAAAACCATTTGGTATGATTTAAATC 91165  
 QY 105 ttccataggtgcacacatactcttcttcaaaactttcagaaagcaaatatataatgaat 164  
 DB 91164 TGCAGAAAATTAATTAATCAATAGAAAGTCTTTTAAAGAGAGAAATATTTAGAAT 91105  
 QY 165 tgancatctactaggaaggatcncagtnaaatcctcncaggtttaaagggtttcacaannn 224  
 DB 91104 TGATCACTATTAGGTAAAGAAATGATTCAAATATATGACGCTTCGATTTGGAAACAC 91045  
 QY 225 agntttgagccacttngagngnancnt 253  
 DB 91044 TATATTGAACCACTTTGGAAATAATAAGT 91016

RESULT 12  
 CCA6246  
 LOCUS  
 DEFINITION Cyanidium caldarium mRNA for glucose-6-phosphate 1-dehydrogenase.  
 ACCESSION AJ006246  
 VERSION AJ006246.1 GI:5734371  
 KEYWORDS g6pdh gene; glucose-6-phosphate 1-dehydrogenase.  
 SOURCE Cyanidium caldarium.  
 ORGANISM Cyanidium caldarium.  
 Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;  
 Porphyridiaceae; Cyanidium.  
 REFERENCE 1 (bases 1 to 1976)  
 AUTHORS Wendt,U.K., Hauschild,R., Lange,C., Pietersma,M., Wenderoth,I. and

von Schaewen,A.  
 Evidence for functional convergence of redox regulation in G6PDH  
 isoforms of cyanobacteria and higher plants  
 Plant Mol. Biol. 40 (3), 487-494 (1999)  
 99364543  
 REFERENCE 2 (bases 1 to 1976)  
 AUTHORS von Schaewen,A.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-MAY-1998) von Schaewen A., Pflanzenphysiologie,  
 Universitaet Osnabrueck, Barbarastr. 11, 49076 Osnabrueck, GERMANY  
 FEATURES  
 Location/Qualifiers  
 source  
 1..1976  
 /organism="Cyanidium caldarium"  
 /db\_xref="taxon:2771"  
 /dev\_stage="vegetative/autotrophic"  
 transit\_peptide 46..258  
 /gene="g6pdh"  
 CDS  
 46..1848  
 /gene="g6pdh"  
 /EC\_number="1.1.1.49"  
 /codon\_start=1  
 /product="glucose-6-phosphate 1-dehydrogenase"  
 /protein\_id="CAB52681.1"  
 /db\_xref="GI:5734372"  
 translation="MWIPLFITFWNTINTSSHLWKTKYTKGWTPLFYFSKVDKFFFA  
 PRLASIQVQGLDNIAPQWNSLETLPTITNERPAIPROHFSHLGDTLDPSLMD  
 VKKQRAKEEPLCIVIGAGDLAKKFTFALFSLYVHLLPKDLIVGARMTQ  
 EEFNRSIMESLTCRVIDGPOCKMDELPKCHYMSGMYDRTEDVRLDQFLNFEQS  
 PNTRVRLYLAVPSQFENVVHVHESGRQGNRIIVMEKPGKDDITSLQLRNS  
 LANCISEDEIRIDHILGKELQNLVLMFLRFANYFEPLNRDHIASIQIVFKNFVE  
 FVLQYRDQNPQSYLSEPGVMNDSTPTFAACVQVDNRNWSGVPLMKAGALDE  
 RKAELRIQFQSVPGFLFSQVYVNSHLPNHLVTVOPDEAIVMRLSKAPGTSRLEE  
 RLNLFTAWEDSKDIPAYERLILDVTHGKSLFIRDELEVAWNIPTPSLKEMEMA  
 QDSWKPIILVDYGGRPIDSYDLAAKYAVQNSEGD"  
 46..1848  
 /gene="g6pdh"  
 mat\_peptide 259..1845  
 /gene="g6pdh"  
 /EC\_number="1.1.1.49"  
 /product="glucose-6-phosphate 1-dehydrogenase"  
 BASE COUNT 623 a 321 c 404 g 628 t  
 ORIGIN

Query Match 21.6%; Score 56.8; DB 8; Length 1976;  
 Best Local Similarity 50.2%; Pred. No. 0.0001;  
 Matches 124; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
 QY 2 aagcacttttgatgttcgctcatgtcttgcagcagtcagaccagaggtatga 61  
 DB 749 AAGTATTTGAAATGTGTATCATCATGTGTGATGAAAGCGAAGAACGACGAGGTGGA 808  
 QY 62 atcgcaataatttgagaagcatttgcttggatgcactttctccataggctgacac 121  
 DB 809 ATCGCATCGTCATGGAACCACTTTGGAAGACATTAATCTTCTATCATAGCTTCGAA 868  
 QY 122 atatcttcttcaaaccttcaggaaagcaaatatataagattgancatctactagaa 181  
 DB 869 ATAGTCTTCGCAATTTGTTCTGAAGATGAATATACCGAATTCATCAGCTTCGAA 928  
 QY 182 ggaatcncagtnaaatcctcncaggtttaaagggtttcacaannnagntttttagccactt 241  
 DB 929 AGGAGTGGTACAAAATTTGATGTTTACGGTTTGCAATTTATCTATTTGAGCCTTTAT 988  
 QY 242 ngagngg 248  
 DB 989 GGAATCG 995

RESULT 13  
 ASU09032  
 LOCUS ASU09032 463 bp mRNA INV 20-JUN-1996

DEFINITION Anastrepha suspensa glucose-6-phosphate dehydrogenase mRNA, partial cds.

ACCESSION U09032

VERSION U09032.1 GI:606621

KEYWORDS Caribbean fruit fly.

SOURCE Anastrepha suspensa

ORGANISM Anastrepha suspensa

REFERENCE 1 (bases 1 to 463)

AUTHORS Soto-Adames,F.N., Robertson,H.M. and Berlocher,S.H.

TITLE Phylogenetic utility of partial DNA sequences of G6PDH at different taxonomic levels in Hexapoda with emphasis on Diptera

JOURNAL Ann. Entomol. Soc. Amer. 87, 723-736 (1994)

AUTHORS Soto-Adames,F.N.

TITLE Direct Submission

JOURNAL Submitted (19-APR-1994) Felipe N. Soto-Adames, Entomology, University of Illinois, 505 S. Goodwin Ave., Urbana, IL 61801, USA

FEATURES

Source

1. .463

/organism="Anastrepha suspensa"

/db\_xref="taxon:28587"

/note="partial sequence from PCR product"

<1..>463

/codon\_start=2

/product="glucose-6-phosphate dehydrogenase"

/protein\_id="AAB02776.1"

/db\_xref="GI:606622"

/translation="LWLFRLNLLPKPKFGYARSKLTVDLRKAKHQYKVPQHEQ

/transl\_except="YVAGGVDQSDFDALRDLRDCDNNRFLYALPSPVQVTV

NIKIDFQSECGWNNVIEKPPGDDASKKLSHSLFHEEQIYRIDHVLG"

BASE COUNT 119 a 107 c 113 g 124 t

ORIGIN

Query Match 21.5%; Score 56.6; DB 30; Length 463;

Best Local Similarity 61.8%; Pred. No. 0.00011;

Matches 89; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

37 agtgcacagaccagaggaatgcataatttggagagccatttgcttgat 96

320 ATTGTCAATCGAATGCGGTGGAATCGGTGATCGAATGCGGTGATCGGCGGTGAC 379

97 gcaatttcctccataggtgcacacaaatttttttcaaaccttcaggaagcaata 156

380 GATGCGCTCTTCAAGAACTGAGCGACCATCTGCTTCACTCTTCCACGAGGCAATC 439

157 tatagaatgancatctacttagga 180

440 TATCGTATGATCATCTATCTGGGA 463

RESULT 14

AF269359

LOCUS AF269359 3015 bp DNA BCT 03-AUG-2000

DEFINITION Staphylococcus epidermidis strain SR1 clone step.1002e12 genomic sequence.

ACCESSION AF269359

VERSION AF269359.1 GI:9664680

KEYWORDS Staphylococcus epidermidis.

SOURCE Staphylococcus epidermidis

ORGANISM Staphylococcus epidermidis

REFERENCE 1 (bases 1 to 3015)

AUTHORS Kimmerly,W.J., Taylor,J.David, Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.

TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis genome

Unpublished

2 (bases 1 to 3015)

Taylor,J.David, Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.

Direct Submission

Submitted (22-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA

genome

Unpublished

2 (bases 1 to 3015)

Taylor,J.David, Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.

Direct Submission

Submitted (19-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA

FEATURES

Source

1. .3015

/organism="Staphylococcus epidermidis"

/strain="SR1"

/db\_xref="taxon:1282"

/clone="step.1002e12"

BASE COUNT 1138 a 384 c 515 g 978 t

ORIGIN

Query Match 20.5%; Score 54; DB 1; Length 3015;

Best Local Similarity 54.5%; Pred. No. 0.00053;

Matches 114; Conservative 0; Mismatches 93; Indels 2; Gaps 1;

45 gaccagagaggtggaatgcataatttggagagccatttgcttgatgcacttc 104

1315 GACGTAAAGGGGTGATGCGATATATTATAGAAACCATTGTTGATGATTTAAATC 1374

105 ttcccataggtgcacacaaattcttcttcaaaccttcaggaagcaaatatagaat 164

1375 T--GCAAAAAATTAATAATCAATAAGAAAGTCTTTTAAAGAAAGAAAAATATTGGAAT 1432

165 tgancatctactaggaaggaatcncagtnaaatcctcaggtttaagggtttcaannn 224

1433 TGATCATATTATTAGGTAAAGAAATGATTCATAATATGACGCTACGATTTTGAACAC 1492

225 agntttgagccaccttngagngnagnt 253

1493 TATATTTGAACCACTTTTGGATAATAAGT 1521

RESULT 15

AF269560/c

LOCUS AF269560 3081 bp DNA BCT 01-AUG-2000

DEFINITION Staphylococcus epidermidis strain SR1 clone step.1010e05 genomic sequence.

ACCESSION AF269560

VERSION AF269560.1 GI:9623454

KEYWORDS Staphylococcus epidermidis.

SOURCE Staphylococcus epidermidis

ORGANISM Staphylococcus epidermidis

REFERENCE 1 (bases 1 to 3081)

AUTHORS Kimmerly,W.J., Taylor,J.David, Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.

Direct Submission

Submitted (22-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA

Drive, Research Triangle Park, North Carolina 27709-3398, USA

FEATURES  
source

1. 3081  
Location/Qualifiers  
/organism="Staphylococcus epidermidis"  
/strain="SR1"  
/db\_xref="taxon:1282"  
/clone="step.1010e05"

BASE COUNT 963 a 528 c 419 g 1171 t  
ORIGIN

Query Match 20.5%; Score 54; DB 1; Length 3081;  
Best Local Similarity 54.5%; Pred. No. 0.00053;  
Matches 114; Conservative 0; Mismatches 93; Indels 2; Gaps 1;  
Qy 45 qaccagaaggatgcataatatttgagaagccatttgcttgatgcacttc 104  
Db 2346 GAACGTAAAGGGTGATGCGGAATAATATAGAAAAACCATTTGGTGATGATTTAAAAATC 2287  
Qy 105 ttccataggtgcacacatatcttcttcaaacctttcaggaaaaagcaaatatagaaat 164  
Db 2286 T--GCAAAAAAATTAAATAATCAATAAGAAAGCTTTTAAAGAGAAAAAATATTGGAAT 2229  
Qy 165 tgancatctactaggaaggaatcncagtnaaatccctncaggtttaagggtttcacaannn 224  
Db 2228 TGATCAGTATTAGGTAAAGAAATGATTCAAAATATTGAACGCTACGATTTTGAACAC 2169  
Qy 225 agntttgagccaccttngagngnaent 253  
Db 2168 TATATTTGAACCCACTTTTGGAAATAATAAGT 2140

Search completed: November 4, 2000, 13:30:39  
Job time: 16888 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:40:19 ; Search time 320.8 Seconds  
(without alignments)  
307.978 Million cell updates/sec

Title: US-09-300-482-4

Perfect score: 263

Sequence: 1 gaagcacttttgatgtgc.....agnngnacntnnnnnganna 263

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.2	22.1	32768	20	X13065
2	51.4	19.5	16592	18	V74364
3	51	19.4	2750	14	Q38808
4	50.6	19.2	2494	21	A26852
5	49.6	18.9	1488	19	V32895
6	48	18.3	2764	19	V65241
7	48	18.3	13121	19	V52259
8	44.8	17.0	1488	19	V56039
9	44.8	17.0	10820	20	X20527
10	42.2	16.0	1482	11	Q05236
11	39.8	15.1	554	20	X21060
12	39.6	15.1	2260	18	T88030
					Enterococcus faeca
					Staphylococcus aur
					G6PD coding sequen
					Essential Staphylo
					Streptococcus pneu
					DNA encoding a S.
					Streptococcus pneu
					Thermostable gluco
					Polynucleotide seq
					Glucose-6-phosphat
					Polynucleotide seq
					Brevibacterium fla

13	39	14.8	780	18	T68126
14	37.4	14.2	1461	16	T30576
15	37.2	14.1	1455	16	T30577
16	37.2	14.1	1467	16	T30575
17	37.2	14.1	1696	13	Q20959
c 18	34.4	13.1	9181	20	X84323
19	33	12.5	1380	21	Z34730
20	33	12.5	1488	21	X34648
21	33	12.5	2042	20	X33972
22	33	12.5	2266	21	Z45852
c 23	33	12.5	9199	20	X84322
24	32.6	12.4	1636	21	Z63255
25	32.4	12.3	600	20	V86185
c 26	32	12.2	87350	18	X83003
c 27	31.6	12.0	1230	10	N90223
28	31.4	11.9	1230025	20	X91990
29	30.4	11.6	1335	20	Z52939
30	30.4	11.6	237326	19	V57903
c 31	30.2	11.5	65921	21	Z89046
32	30	11.4	1380	21	Z34650
33	29.8	11.3	1458	16	T30574
34	29.8	11.3	1963	13	Q24526
35	29.8	11.3	11901	20	X02998
36	29.6	11.3	1484	20	X89755
37	29.6	11.3	1517	14	Q47166
38	29.6	11.3	1517	14	Q47167
39	29.6	11.3	1517	19	V23923
40	29.6	11.3	1517	19	V24099
41	29.6	11.3	1517	19	V25817
42	29.6	11.3	1517	19	V25818
43	29.6	11.3	1517	19	V09855
44	29.6	11.3	1517	19	V09856
45	29.6	11.3	1517	19	V12162

#### ALIGNMENTS

RESULT 1  
X13065  
ID X13065 standard; DNA; 32768 BP.  
XX  
AC X13065;  
XX  
DT 19-MAR-1999 (first entry)  
XX  
DE Enterococcus faecalis genome contig SEQ ID NO:128.  
XX  
KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
KW vaccine; attenuation; computer readable medium; ds.  
XX  
OS Enterococcus faecalis.  
XX  
PN W09850555-A2.  
XX  
PD 12-NOV-1998.  
XX  
PF 04-MAY-1998; 98WO-US08985.  
XX  
PR 14-NOV-1997; 97US-0066009.  
PR 06-MAY-1997; 97US-0044031.  
PR 16-MAY-1997; 97US-0046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Barash SC, Dillon PJ, Kunsch CA;  
XX  
DR WPI; 1999-045171/04.  
XX  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.

XX PS Claim 1; Page 767-783; 2084pp; English.

XX CC A computer readable medium has been developed which has recorded on it

CC CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC CC X12938 to X13919 represent these nucleotide sequences which are primary

CC CC nucleotide sequences, also known as contigs. The computer-based system

CC CC can identify fragments of the Enterococcus faecalis genome with

CC CC commercial importance. The products can be used to detect the presence

CC CC of Enterococcus faecalis in samples. They can also be used for

CC CC diagnosing Enterococcal infection in an animal and monitoring

CC CC progression of disease, and for identifying agents which can be used to

CC CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC CC another related organism, in vivo or in vitro. In particular the

CC CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC CC can be used in vaccines to prevent or attenuate an Enterococcal

CC CC infection.

XX SQ Sequence 32768 BP; 9529 A; 6883 C; 5611 G; 10720 T; 25 other;

Query Match 22.1%; Score 58.2; DB 20; Length 32768;

Best Local Similarity 51.2%; Pred. No. 9.5e-08;

Matches 108; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 53 agggatggaatgcataatttgagagccatttggttgatgcactttctccata 112

Db 19293 aaggcttgatcgctcaattatcgaaaagccatttggttgatgcacttgctgcacg 19352

QY 113 ggcgcacacatactctttcaaaccttcagagaaagcaaatatagaaatgancatc 172

Db 19353 aattaaatgaagaattcgctgcattcttcgcagcaagataatttccgaattgaccatt 19412

QY 173 tactaggaaggaaatcncagtnaaatcctnccaggttttaagggtttccaaannnagntttg 232

Db 19413 acttaggcaagaatgatccaaatatttcagccattcggtttgccaataaattttg 19472

QY 233 agccaccttgagngnagntnnnganna 263

Db 19473 aatcacaatggacaatcgattacattgataa 19503

RESULT 2

V74364

ID V74364 standard; DNA; 16592 BP.

XX AC V74364;

XX DE 16-MAR-1999 (first entry)

XX DE Staphylococcus aureus contig SEQ ID #53.

XX KW Computer readable medium; vaccine; S. aureus infection; immunodetection;

XX KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

XX KW skin infection; surgical wound infection; scalded skin syndrome;

XX KW toxic shock syndrome; ds.

XX OS Staphylococcus aureus.

XX FH Key Location/Qualifiers

FT FT misc\_feature 121..180

FT FT /\*tag= a

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 1921..1980

FT FT misc\_feature

FT FT /\*tag= b

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 3721..3780

FT FT misc\_feature

FT FT /\*tag= c

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 5521..5580

FT FT misc\_feature

FT FT /\*tag= d

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 7321..7380

FT FT misc\_feature

FT FT /\*tag= e

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 9121..9180

FT FT misc\_feature

FT FT /\*tag= f

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 10921..10980

FT FT misc\_feature

FT FT /\*tag= g

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 12721..12780

FT FT misc\_feature

FT FT /\*tag= h

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 14521..14580

FT FT misc\_feature

FT FT /\*tag= i

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 16321..16380

FT FT misc\_feature

FT FT /\*tag= j

FT FT /note= "these bases represent a line of missing text in

FT FT the sequence listing in the specification. They

FT FT are included to maintain the nucleotide numbering

FT FT given in the specification for this DNA sequence"

FT FT 16321..16380

PN EP786519-A2.

XX 30-JUL-1997.

XX 07-JAN-1997; 97EP-0100117.

XX 05-JAN-1996; 96US-0009861.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

PI Rosen CA;

XX WPI; 1997-374922/35.

DR Polynucleotide(s) and proteins derived from Staphylococcus aureus

PT stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

XX Claim 1; Page 403-413; 3271pp; English.

PS This sequence represents one of 5191 Staphylococcus aureus DNA sequences

CC of the invention. The DNA sequences are recorded on a computer readable

CC medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using



CC includes a method for evaluating an agent for activity on the essential  
CC genes. Also included in the invention is the production of an  
CC antibacterial agent. The antibacterial agents of the invention are used  
CC to treat or prevent bacterial infections, particularly where caused by  
CC *Streptococcus aureus*. The antibacterial agents are unlikely to be  
CC affected by known resistance mechanisms.

XX Sequence 2494 BP; 809 A; 396 C; 389 G; 880 T; 20 other;

Query Match 19.28; Score 50.6; DB 21; Length 2494;  
Best Local Similarity 50.08; Pred. No. 7.5e-06;  
Matches 101; Conservative 2; Mismatches 99; Indels 0; Gaps 0;  
QY 52 aaggatggaatgcataatattgagagccatttggcttggatgcactttctcccat 111  
DB 1425 AAGAGTAAACGGCTTGTATCGAAACCATTCGGTAGTGATTTAAATCAGCGGAA 1366  
QY 112 aggtgacacataatcttcttcaactttcaggagaaagcaaatatagaaattgancat 171  
DB 1365 GCATTAAACAATCAATTCGTAATCAATTAAGAAAGAAATTTAYCGTATGACAC 1306  
QY 172 ctactaggaagaatcncagtnaaatcctncaggttttaagggtttcaaaannnagntttt 231  
DB 1305 TATTAGGAAAGACATCGTTCAAAATATCGAGGTATTACGTTTTCGATCGATGTTT 1246  
QY 232 gagccaccttngagngnacnt 253  
DB 1245 GAACCATTTAGCAATTAACAAAT 1224

RESULT 5  
V32895  
ID V32895 standard; DNA; 1488 BP.  
XX  
AC V32895;  
XX  
DT 26-OCT-1998 (first entry)  
DE Streptococcus pneumoniae glucose 6-phosphate dehydrogenase encoding DNA.  
XX  
KW Glucose 6-phosphate dehydrogenase; G6PD; meningitis; ds.  
XX Streptococcus pneumoniae.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1488  
FT /\*tag= a  
FT /\*product= "Streptococcus pneumoniae G6PD"  
XX  
PN EF854187-A2.  
XX  
PD 22-JUL-1998.  
XX  
PF 16-JAN-1998; 98EP-0300315.  
XX  
PR 21-JAN-1997; 97US-0035072.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PI Burnham MKR;  
XX  
DR WPI; 1998-379058/33.  
DR P-PSDB; W49070.  
XX  
PT New nucleic acid encoding glucose-6-phosphate dehydrogenase -  
PT isolated from Streptococcus pneumoniae, used to treat, prevent and  
PT diagnose infections, especially meningitis  
XX  
PS Claim 4; Fig 1-1B; 30pp; English.  
XX  
CC The present sequence represents the Streptococcus pneumoniae glucose  
CC 6-phosphate dehydrogenase (G6PD) encoding DNA sequence. The

CC S. pneumoniae G6PD DNA and protein are claimed to be useful for  
CC treating, preventing and diagnosing S. pneumoniae infections such  
CC as meningitis. Primers (V32896-V32897) and probes based on the G6PD  
CC DNA sequence are claimed to be useful for determining the stage and  
CC type of infection, and also for identifying and isolating related  
CC sequences.

XX Sequence 1488 BP; 449 A; 328 C; 293 G; 418 T; 0 other;

Query Match 18.9%; Score 49.6; DB 19; Length 1488;  
Best Local Similarity 48.6%; Pred. No. 1.3e-05;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
QY 52 aaggatggaatgcataatattgagagccatttggcttggatgcactttctcccat 111  
DB 412 aaagggttttgagcgttgatcgttgaaaacccatttggtagagattacgaactgcaagc 471  
QY 112 aggtgacacataatcttcttcaactttcaggagaaagcaaatatagaaattgancat 171  
DB 472 aegttgaatgacgaactcctagcaacatttgacgaagaaacaaatttccgtattgacat 531  
QY 172 ctactaggaagaatcncagtnaaatcctncaggttttaagggtttcaaaannnagntttt 231  
DB 532 tatcttggttaaggaaatgatccaaagcatcttgcagttcgttgcgaactgtatttt 591  
QY 232 gagccaccttngagngnacntnnnnnanna 263  
DB 592 gaaacggttggaaacgaagattttatcgacaa 623

RESULT 6  
V65241/c  
ID V65241 standard; DNA; 2764 BP.  
XX  
AC V65241;  
XX  
DT 24-DEC-1998 (first entry)  
DE DNA encoding a S. pneumoniae protein having function in cell division.  
XX Streptococcus pneumoniae protein; recombinant; gene expression;  
KW DNA chip; virulence; antibody; infection; detection; treatment; ss.  
XX Streptococcus pneumoniae.  
XX  
PN WO9826072-A1.  
XX  
PD 18-JUN-1998.  
XX  
PF 09-DEC-1997; 97WO-US22578.  
XX  
PR 13-DEC-1996; 96US-0036281.  
XX  
PA (ELIL ) LILLY & CO ELI.  
XX  
PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;  
PI Mills BJ, Norris FH, Peery RB, Rostek PR;  
PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;  
PI Young Bellido ML;  
XX  
DR WPI; 1998-348529/30.  
XX  
PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips  
PT for evaluating gene expression, and identification of virulence  
XX genes  
XX  
PS Claim 1; Page 94; 333pp; English.  
XX  
CC This is the DNA sequence of a S. pneumoniae protein having function in  
CC cell division. The invention provides DNA sequences (V65201 to V65304)  
CC from the S. pneumoniae genome and corresponding protein sequences  
CC (W80605 to W80728). A recombinant host containing a vector comprising any



CC of the above nucleic acids can be used for the recombinant expression of  
 CC the protein sequences. The invention also provides a DNA chip having  
 CC arrayed on it at least 15 base pair fragment of any one or more of these  
 CC DNA sequences. The DNA chip can be used methods for evaluating gene  
 CC expression in *S. pneumoniae* and for identifying virulence genes in  
 CC *S. pneumoniae*. Antibodies that selectively bind to the above proteins or  
 CC peptide fragments can be used to treat *S. pneumoniae* infection. The  
 CC antibodies can also be used to detect *S. pneumoniae* cells.  
 XX  
 SQ Sequence 2764 BP; 829 A; 515 C; 646 G; 774 T; 0 other;

Query Match 18.3%; Score 48; DB 19; Length 2764;  
 Best Local Similarity 51.3%; Pred. No. 4.6e-05;  
 Matches 99; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
 QY 52 aaggatggaatgcacataattttgagaagccatttggttgatgcactttcttccat 111  
 Db 2743 AAAGGTTTTTGAGCGCTTGATCGTTGAAAACCAATTTGTACAGATTACGCAACTGCAAC 2684  
 QY 112 aggtgacacacatactcttcttcaaaccttcaggaaagcaaatatataagaattgancat 171  
 Db 2683 AGTTGATGACGAACTCCTAGCAACATTTGACGAAGAACAAATTTTCCTATCGACCAT 2624  
 QY 172 ctactaggaagaatcncagtnaaatcctcncaggttttaagggtttcaaaannnagntttt 231  
 Db 2623 TATCTTGTGAAGGAAATGATCAAGCATCTTTGCAGTTCGCTTTCGAAACTTGATTTT 2564  
 QY 232 gaggcaccttnga 244  
 Db 2563 GAAACGTTTGGG 2551

RESULT 7  
 V52259  
 ID V52259 standard; DNA; 13121 BP.  
 XX  
 AC V52259;  
 XX  
 DT 23-OCT-1998 (first entry)  
 XX Streptococcus pneumoniae genome fragment SEQ ID NO:126.  
 XX Streptococcus pneumoniae; *S. pneumoniae*; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX Streptococcus pneumoniae.  
 OS  
 XX W09818931-A2.  
 XX  
 XX 07-MAY-1998.  
 XX  
 XX 30-OCT-1997; 97WO-US19588.  
 XX  
 XX 31-OCT-1996; 96US-0029960.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Barash SC, Choi GH, Dillion PU, Dougherty BA, Fannon M;  
 PI Kunsch CA, Rosen CA;  
 XX WPI; 1998-272225/24.  
 XX  
 XX Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 XX  
 PS Claim 1; Page 869-876; 1409pp; English.  
 XX  
 XX The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical

CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the *S. pneumoniae*  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridize to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the *S. pneumoniae* genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC *S. pneumoniae* genome of commercial importance, or expression/modulating  
 CC fragments of the *S. pneumoniae* genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for *S. pneumoniae*.  
 XX  
 SQ Sequence 13121 BP; 3698 A; 3021 C; 2439 G; 3963 T; 0 other;

Query Match 18.3%; Score 48; DB 19; Length 13121;  
 Best Local Similarity 48.1%; Pred. No. 7.5e-05;  
 Matches 102; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
 QY 52 aaggatggaatgcacataattttgagaagccatttggttgatgcactttcttccat 111  
 Db 3492 aaaggttttgagcgcgttgatcgcttgaaacacatttggtacagattacgcaactgcaagc 3551  
 QY 112 aggtgacacacatactcttcttcaaaccttcaggaaagcaaatatataagaattgancat 171  
 Db 3552 aagttgaatgacgaactcctagcaacatttgacgaagaacaaatttcgctatcgaccat 3611  
 QY 172 ctactaggaagaatcncagtnaaatcctcncaggttttaagggtttcaaaannnagntttt 231  
 Db 3612 tatctgtgtaagaaatgatccaaagcatcttgcggttcgctttgcaaaactgtatttt 3671  
 QY 232 gaggcaccttngagngnagcnnnnnnnanna 263  
 Db 3672 gaaacgtttggacaagagattttatcgacaa 3703

RESULT 8  
 V56039  
 ID V56039 standard; DNA; 1488 BP.  
 XX  
 AC V56039;  
 XX  
 DT 09-DEC-1998 (first entry)  
 XX Thermostable glucose-6-phosphate dehydrogenase encoding DNA.  
 DE  
 XX Glucose-6-phosphate dehydrogenase; thermostable; heat-resistant;  
 KW enzyme; recombinant; high stability; ds.  
 XX  
 XX Pedicoccus pentosaceus.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1488  
 FT /tag= a  
 FT /product= "glucose-6-phosphate dehydrogenase"  
 XX JP10225293-A.  
 XX  
 XX 25-AUG-1998.  
 PD  
 XX  
 PF 13-FEB-1997; 97JP-0029345.  
 XX  
 PR 13-FEB-1997; 97JP-0029345.  
 XX  
 PA (KIRK ) KIKKOMAN CORP.

```
XX WPI; 1998-513903/44.
DR P-PSDB; W80567.
XX Heat-resistant glucose-6-phosphate dehydrogenase gene - useful for
PT efficiently preparing heat-resistant glucose-6-phosphate
PT dehydrogenase of high stability
XX
XX Claim 1; Pages 5-6; 7pp; Japanese.
XX
XX This DNA encodes a heat-resistant glucose-6-phosphate dehydrogenase. A
CC transformant or a transducent containing a recombinant DNA in which the
CC heat-resistant (thermostable) glucose-6-phosphate dehydrogenase gene is
CC inserted to a vector DNA, can be cultured to produce the enzyme
CC recombinantly. The method can prepare heat-resistant glucose-6-phosphate
CC dehydrogenase of high stability efficiently.
XX
XX Sequence 1488 BP; 496 A; 249 C; 308 G; 435 T; 0 other;
SQ
Query Match 17.0%; Score 44.8; DB 19; Length 1488;
Best Local Similarity 47.2%; Pred. No. 0.00033;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 52 aaggatggaatcgataataatttgagagcatttggttgatgcacttttcccat 111
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 112 aggtgacacatactcttttcaaaccttcaggaaagcaaatatagaaattgancat 171
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 490 gaatgaatgattcaatttcggttccttaagaagaacagatttttagaattgatcat 549
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 172 ctactaggagaagaatcncagtnaaatcctncaggttttaagggtttcaaaannnagntttt 231
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 550 tatttagttaaggaaacccgacatcagtcacattgctgattacggtttgggaacgcatgttc 609
QY 232 gagccacttngagnnagntnnnnnganna 263
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 610 aactcgtttggaataagaagaacatttgataa 641
RESULT 9
X20527/c
ID X20527 standard; DNA; 10820 BP.
XX
AC X20527;
XX
DT 05-MAY-1999 (first entry)
XX
DE Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX
OS Treponema pallidum.
XX
PN W09859034-A2.
XX
PD 30-DEC-1998.
XX
XX 23-JUN-1998; 98WO-US13041.
XX
XX 24-JUN-1997; 97US-0050667.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fraser CM;
XX
XX WPI; 1999-081273/07.
XX
XX New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
```

```
XX Claim 1; Page 343-349; 1150pp; English.
XX
XX X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
XX Sequence 10820 BP; 2418 A; 2982 C; 2962 G; 2450 T; 8 other;
SQ
Query Match 17.0%; Score 44.8; DB 20; Length 10820;
Best Local Similarity 48.0%; Pred. No. 0.00062;
Matches 94; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 68 taatattgagaagccatttggttgatgcacttttcccatagggtgacacaaatc 127
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 ttctttcaaaccttcagaaagcaaatatagaaattgancatctactaggaaggaatc 187
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6607 TGGTGCCCATCTTTCAGGAAACCAACCTATCGCATCATCTATCTCGGTAGGAAA 6548
QY 188 ncagtnaaatcctncaggttttaagggttttcaaaannnagnttttgagccacttngagnn 247
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6547 CGGTCCAAACATCCTGGTCACCTGCTTTGCCAATCCCTTTTCGAGCGCCACATGGAAC 6488
QY 248 gnacntnnnnnganna 263
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6487 GGACCCATATCGATTA 6472
RESULT 10
Q05236
ID Q05236 standard; DNA; 1482 BP.
XX
AC Q05236;
XX
DT 12-NOV-1990 (first entry)
XX
DE Glucose-6-phosphate dehydrogenase gene.
XX
KW Glucose-6-phosphate dehydrogenase; creatine kinase; ss.
XX
OS Bacillus.
XX
FH Key Location/Qualifiers
FT CDS 1..1482
FT FT /*tag= a
XX
XX DE3942129-A.
XX
XX 05-JUL-1990.
XX
XX 20-DEC-1989; 89DE-3942129.
XX
XX 29-DEC-1988; 88JP-0334629.
XX
XX (TOXN ) TOYO JOZO KK.
XX
XX Sagai H, Hattori K, Takahashi M;
XX
XX WPI; 1990-211190/28.
XX
XX P-PSDB; R05798.
XX
XX New glucose-6-phosphate dehydrogenase gene from Bacillus species -
XX used for expressing stable, very pure enzyme, useful in assay of
XX creatine kinase.
XX
XX Disclosure; ; p; German.
```

The protein produced, on culturing host cells transformed with a vector contg. this sequence, is useful as a reagent for assaying creatine kinase. It has improved long term and thermal stability over known forms of the enzyme.

Query Match	16.0%;	Score 42.2;	DB 11;	Length 1482;	
Best Local Similarity	46.1%;	Pred. N.0.002;			
Matches 101;	Conservative	0;	Mismatches 118;	Indels 0;	Gaps
QY	45	gaccacgaaggatgaatcgcaataattttgagaagccatttggttgatgcacttc	104		
Db	411	ggccacaaacggatgacgcgattagtcgatgataaaacccgttcggtcatgattacaag	470		
QY	105	ttcccataggtgacacaaatctctttccaaactttcagaaaagcgaataatagaaat	164		
Db	471	cgcacaaaactaaatgaagaaatcgccaatcgttttcagaggagcaaatttccgtat	530		
QY	165	tgancattactagggaagaaatcncagtnaaactctncaggttttaagggtttcaaan	224		
Db	531	tgaccattatctcgcaagaagaatggtgcacaacatcgaggtcattcgcttttcgcaagc	590		
QY	225	agntttgagccaccttngagngnacntnnnnnganna	263		
Db	591	cattttgcaccgctctggaataaacgcgtttattgcac	629		

RESULT	11
X21060	
ID	X21060 standard; DNA; 554 BP.
XX	
AC	X21060;
XX	
DT	05-MAY-1999 (first entry)
XX	
DE	Polynucleotide sequence from the genome of <i>Treponema pallidum</i> .
XX	
KW	<i>Treponema pallidum</i> infection; syphilis; <i>Borrelia</i> infection; animal;
KW	enzyme production; ds.
XX	
OS	<i>Treponema pallidum</i> .
XX	
PN	W09859034-A2.
XX	
PD	30-DEC-1998.
XX	
PF	23-JUN-1998; 98WO-US13041.
XX	
PR	24-JUN-1997; 97US-0050667.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Fraser CM;
XX	
WPI	1999-081273/07.
XX	
PT	New isolated <i>Treponema pallidum</i> nucleic acids - used to develop
PT	products for the detection, diagnosis, characterisation, prevention
PT	and therapy of <i>T. pallidum</i> infections, particularly syphilis
XX	
PPS	Claim 1; Page 1045; 1150pp; English.
XX	
CCC	X20500-21243 represent polynucleotide sequences from the genome of
CCC	<i>Treponema pallidum</i> . The sequences can be used for detection,
CCC	diagnosis, characterisation, prevention and therapy for <i>T. pallidum</i>
CCC	infections, particularly syphilis. They can also be used for detection
CCC	diseases related to <i>Borrelia</i> infections in animals, and for the
CCC	production of biosynthetic products such as enzymes.
XX	
SQ	Sequence 554 BP; 141 A; 159 C; 126 G; 127 T; 1 other;

[illegible]

RESULT	12	
T88030		
ID	T88030	standard; DNA; 2260 BP.
XX	XX	
XX	AC	
T88030;		
XX	XX	
XX	XX	
DT	DT	
XX	XX	
DE	DE	
XX	XX	Brevibacterium flavum glucose-6-phosphate dehydrogenase DNA.
XX	XX	Brevibacterium flavum; glucose-6-phosphate dehydrogenase;
KW	KW	recombinant production; coryneform; bacteria; bacterium; ds.
KW	KW	
OS	OS	Brevibacterium flavum.
XX	XX	
Key	Location/Qualifiers	
CDS	629..2083	
FT	/*tag= a	
FT	/product= glucose-6-phosphate_dehydrogenase	
FT		
PN	PN	
XX	XX	JP09224661-A.
PD	PD	
XX	XX	02-SEP-1997.
XX	XX	
PF	PF	23-FEB-1996; 96JP-0036345.
XX	XX	
PR	PR	23-FEB-1996; 96JP-0036345.
XX	XX	
PA	PA	(MITU ) MITSUBISHI CHEM CORP.
XX	XX	
DR	DR	WPI; 1997-484096/45.
DR	DR	P-PSDB; W27612.
XX	XX	
PPT	PPT	Glucose-6-phosphate dehydrogenase - which can be recombinantly
PT	PT	produced by transforming coryneform bacteria with DNA molecule
PT	PT	encoding it
XX	XX	
PS	PS	Claim 1; Pages 6-8; 9pp; Japanese.
XX	XX	
CC	CC	The present sequence encodes the Brevibacterium flavum JM-233
CC	CC	glucose-6-phosphate dehydrogenase (G-6-PD). The G-6-PD can be
CC	CC	recombinantly produced by transforming coryneform bacteria with
CC	CC	DNA molecule encoding it.
XX	XX	
SQ	SQ	Sequence 2260 BP; 512 A; 659 C; 592 G; 497 T; 0 other;

Query Match 15.1%; Score 39.6; DB 18; Length 2260;  
Best Local Similarity 45.4%; pred. No. 0.013;  
Matches 99; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

2.



PS Disclosure; Page 60; 122pp; English.

G6PDH genes (T30574-77) respectively code for the glucose-6-phosphate dehydrogenases (R95961-64) of *Leuconostoc mesenteroides* ATCC 12291, *Leuconostoc citreum* NCIMB 3351, *Leuconostoc lactis* NCDO 546 and *Leuconostoc dextranum* ATCC 19255. The enzymes may be mutated to increase stability, or to improve or modulate activity, pref. by deletion or substn. of 1 or more lysine residues, or introduction of 1 or more cysteine residues. The mutant enzymes are prep'd. by mutating the G6PDH gene and expression in host cells. They are used partic. to form conjugates for use in homogeneous immunoassays for analytes.

Sequence 1461 BP; 395 A; 284 C; 324 G; 377 T; 81 other;

Query Match	Score 37.4;	DB 16;	Length 1461;
Best Local Similarity	14.2%;		
	44.7%;	Pred. No. 0.052;	

[illegible]

**RESULT** 15

T30577  
ID T30577 standard; DNA; 1455 BP.

T30577;

DT 14-SEP-1996 (first entry)

*Leuconostoc dextranicum* glucose-6-phosphate dehydrogenase gene.

Glucose-6-phosphate dehydrogenase; G6PDH; immunoassay; ss.

.....  
OS  
.....  
Leuconostoc dextranicum strain ATCC 19255.

WO9424559-A2.

27-OCT-1994.

07-APR-1994; 94WO-US03437.

08-APR-1993; 93US-0044857.

BEHRINGWERKE AG.

A (GEMV ) GENENCOR INT INC.

A (SYNT) SYNTAX USA INC.

Barnett CC, Becker MJ, Bott RR, Caldwell RM, Goodman TC;  
Jakobovits EB, Yarns MT, ...

WPI; 1995-006211/01.

R  
P-PSDB; R95964.  
WET, 1393-00621.

New mutant glucose-6-phosphate dehydrogenase enzymes - used partic. to form conjugates for use in homogeneous immunoassays for analytes

Disclosure; Page 62-63; 122pp; English.

XX

CC G6PDH genes (T30574-77) respectively code for the glucose-6-phosphate  
CC dehydrogenases (R95961-61) of *Leuconostoc mesenteroides* ATCC 12291,  
CC *Leuconostoc citreum* NCIM 3351, *Leuconostoc lactis* NCD0 546 and  
CC *Leuconostoc dextranicum* ATCC 19255. The enzymes may be mutated to  
CC increase stability, or to improve or modulate activity, pref. by  
CC deletion or substn. of 1 or more lysine residues, or introduction of  
CC 1 or more cysteine residues. The mutant enzymes are prepd. by mutating  
CC the G6PDH gene and expression in host cells. They are used partic. to  
CC form conjugates for use in homogeneous immunoassays for analytes.  
XX  
SQ Sequence 1455 BP; 452 A; 275 C; 318 G; 410 T; 0 other;

Sequence 1455 BP; 452 A; 275 C; 318 G; 410 T; 0 other;

Query Match 14.1%; Score 37.2; DB 16; Length 1455;  
Best Local Similarity 44.9%; Pred. No. 0.059;  
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

50	agaaggatggaatcgacataattttgagaagccatttgcctttgatgcactttttccc	109
Qy		
110	ataggctgacacaatatcttttcaaaacttcaggaaaaagcaaatatatagaattganc	169
Qy		
120	ataggctgacacaatatcttttcaaaacttcaggaaaaagcaaatatatagaattganc	169
Qy		
130	ataggctgacacaatatcttttcaaaacttcaggaaaaagcaaatatatagaattganc	169
Qy		
140	ataggctgacacaatatcttttcaaaacttcaggaaaaagcaaatatatagaattganc	169
Qy		
150	ataggctgacacaatatcttttcaaaacttcaggaaaaagcaaatatatagaattganc	169
Qy		
160	ataggctgacacaatatcttttcaaaacttcaggaaaaagcaaatatatagaattganc	169
Qy		
170	atctactaggaaaggaatcncagtnaaactcctncaggttttaagggttttcaaanngntt	229
Qy		
180	atctactaggaaaggaatcncagtnaaactcctncaggttttaagggttttcaaanngntt	229
Qy		
190	atctactaggaaaggaatcncagtnaaactcctncaggttttaagggttttcaaanngntt	229
Qy		
200	atctactaggaaaggaatcncagtnaaactcctncaggttttaagggttttcaaanngntt	229
Qy		
210	atctactaggaaaggaatcncagtnaaactcctncaggttttaagggttttcaaanngntt	229
Qy		
220	atctactaggaaaggaatcncagtnaaactcctncaggttttaagggttttcaaanngntt	229
Qy		
230	tttagccactcttgagngnagctnnnnnganna	263
Qy		
240	tttagccactcttgagngnagctnnnnnganna	263
Qy		
250	tttagccactcttgagngnagctnnnnnganna	263
Qy		
260	tttagccactcttgagngnagctnnnnnganna	263
Qy		
270	tttagccactcttgagngnagctnnnnnganna	263
Qy		
280	tttagccactcttgagngnagctnnnnnganna	263
Qy		
290	tttagccactcttgagngnagctnnnnnganna	263
Qy		
300	tttagccactcttgagngnagctnnnnnganna	263
Qy		
310	tttagccactcttgagngnagctnnnnnganna	263
Qy		
320	tttagccactcttgagngnagctnnnnnganna	263
Qy		
330	tttagccactcttgagngnagctnnnnnganna	263
Qy		
340	tttagccactcttgagngnagctnnnnnganna	263
Qy		
350	tttagccactcttgagngnagctnnnnnganna	263
Qy		
360	tttagccactcttgagngnagctnnnnnganna	263
Qy		
370	tttagccactcttgagngnagctnnnnnganna	263
Qy		
380	tttagccactcttgagngnagctnnnnnganna	263
Qy		
390	tttagccactcttgagngnagctnnnnnganna	263
Qy		
400	tttagccactcttgagngnagctnnnnnganna	263
Qy		
410	tttagccactcttgagngnagctnnnnnganna	263
Qy		
420	tttagccactcttgagngnagctnnnnnganna	263
Qy		
430	tttagccactcttgagngnagctnnnnnganna	263
Qy		
440	tttagccactcttgagngnagctnnnnnganna	263
Qy		
450	tttagccactcttgagngnagctnnnnnganna	263
Qy		
460	tttagccactcttgagngnagctnnnnnganna	263
Qy		
470	tttagccactcttgagngnagctnnnnnganna	263
Qy		
480	tttagccactcttgagngnagctnnnnnganna	263
Qy		
490	tttagccactcttgagngnagctnnnnnganna	263
Qy		
500	tttagccactcttgagngnagctnnnnnganna	263
Qy		
510	tttagccactcttgagngnagctnnnnnganna	263
Qy		
520	tttagccactcttgagngnagctnnnnnganna	263
Qy		
530	tttagccactcttgagngnagctnnnnnganna	263
Qy		
540	tttagccactcttgagngnagctnnnnnganna	263
Qy		
550	tttagccactcttgagngnagctnnnnnganna	263
Qy		
560	tttagccactcttgagngnagctnnnnnganna	263
Qy		
570	tttagccactcttgagngnagctnnnnnganna	263
Qy		
580	tttagccactcttgagngnagctnnnnnganna	263
Qy		
590	tttagccactcttgagngnagctnnnnnganna	263
Qy		
600	tttagccactcttgagngnagctnnnnnganna	263
Qy		
610	tttagccactcttgagngnagctnnnnnganna	263
Qy		
620	tttagccactcttgagngnagctnnnnnganna	263
Qy		
630		

Search completed: November 4, 2000, 13:40:30  
Job time: 16157 sec

100

100







Best Local Similarity 44.7%; Pred. No. 0.0059;  
Matches 98; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 45 gaccagaaggatgcataataatttgagaagccatttgcttgatgcacttc 104  
Db 411 GGCACAACTGGTTACACCGTTTGATGATCGAAAAGCCATTGGGACATCATACGAAC 470  
QY 105 ttccatagctgacacataatctcttcaaaccttcaggaaaagcaaatatagaat 164  
Db 471 AGCTGAAAGTTGCAAAACGAATTGGAAAACGCCCTTTGATGATGACCAATTGTTCCGTAT 530  
QY 165 tgancatctactaggaagaatcncagtnaaaatcctncaggtttaagggtttcaaan 224  
Db 531 TGACCACACTACCTTGGTAAGGAATGGTCCAAAATATTGGCGCTTTGGCTTTGTAACCC 590  
QY 225 agnttttagccacctngagngnnaacntnnnnnganna 263  
Db 591 AATCTTTGATGAGCGCTGGAAACAGGACTACATCAAGAA 629

## RESULT 4

US-08-445-464C-5  
; Sequence 5, Application US/08445464C  
; Patent No. 6090567  
; GENERAL INFORMATION:  
; APPLICANT: Jakobovits, Edward B.  
; APPLICANT: Silen, Joy L.  
; APPLICANT: Levy, Mark J.  
; APPLICANT: Goodman, Thomas C.  
; APPLICANT: Becker, Martin  
; APPLICANT: Ullman, Edwin F.  
; APPLICANT: Caldwell, Robert M.  
; APPLICANT: Bott, Richard R.  
; APPLICANT: Barnett, Christopher C.  
; TITLE OF INVENTION: Homogeneous Immunoassays Using Mutant  
; TITLE OF INVENTION: Glucose-6-Phosphate Dehydrogenases  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dade Behring Inc.  
; STREET: 1717 Deerfield Road  
; CITY: Deerfield  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60015

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,464C  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/044,857  
; FILING DATE: 08-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ruzsala, Lois K.  
; REGISTRATION NUMBER: 39,074  
; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1  
; TELEPHONE: (847) 267-5364  
; TELEFAX: (847) 267-5376

INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1461 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-445-464C-5

Query Match 14.2%; Score 37.4; DB 5; Length 1461;  
Best Local Similarity 44.7%; Pred. No. 0.0059;  
Matches 98; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 45 gaccagaaggatgcataataatttgagaagccatttgcttgatgcacttc 104  
Db 411 GGCACAACTGGTTACACCGTTTGATGATCGAAAAGCCATTGGGACATCATACGAAC 470  
QY 105 ttccatagctgacacataatctcttcaaaccttcaggaaaagcaaatatagaat 164  
Db 471 AGCTGAAAGTTGCAAAACGAATTGGAAAACGCCCTTTGATGATGACCAATTGTTCCGTAT 530  
QY 165 tgancatctactaggaagaatcncagtnaaaatcctncaggtttaagggtttcaaan 224  
Db 531 TGACCACACTACCTTGGTAAGGAATGGTCCAAAATATTGGCGCTTTGGCTTTGTAACCC 590  
QY 225 agnttttagccacctngagngnnaacntnnnnnganna 263  
Db 591 AATCTTTGATGAGCGCTGGAAACAGGACTACATCAAGAA 629

## RESULT 5

PCT-US94-03437-5  
; Sequence 5, Application PC/TUS9403437  
; GENERAL INFORMATION:  
; APPLICANT:  
; APPLICANT:  
; TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT  
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES  
; NUMBER OF SEQUENCES: 124  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/03437  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1461 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA genomic  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Leuconostoc lactis  
; STRAIN: NCDO 546

PCT-US94-03437-5

Query Match 14.2%; Score 37.4; DB 6; Length 1461;  
Best Local Similarity 44.7%; Pred. No. 0.0059;  
Matches 98; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 45 gaccagaaggatgcataataatttgagaagccatttgcttgatgcacttc 104  
Db 411 GGCACAACTGGTTACACCGTTTGATGATCGAAAAGCCATTGGGACATCATACGAAC 470  
QY 105 ttccatagctgacacataatctcttcaaaccttcaggaaaagcaaatatagaat 164  
Db 471 AGCTGAAAGTTGCAAAACGAATTGGAAAACGCCCTTTGATGATGACCAATTGTTCCGTAT 530  
QY 165 tgancatctactaggaagaatcncagtnaaaatcctncaggtttaagggtttcaaan 224  
Db 531 TGACCACACTACCTTGGTAAGGAATGGTCCAAAATATTGGCGCTTTGGCTTTGTAACCC 590  
QY 225 agnttttagccacctngagngnnaacntnnnnnganna 263



NAME: Ruzsala, Lois K.  
REGISTRATION NUMBER: 39,074  
REFERENCE/DOCKET NUMBER: BEH-7261 DIV 2  
TELEPHONE: (847) 267-5364  
TELEFAX: (847) 267-5376  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1571 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-445-463B-3

Query Match 14.1%; Score 37.2; DB 5; Length 1571;  
Best Local Similarity 44.9%; Pred. No. 0.007;  
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 50 agaagggatggaatcgcaataatttgaagccatttgcttggatgcactttcttccc 109  
Db 501 ATAGTGGTACACCGTTTGATGATTGAAAGCCCTTTGGTACATCATACGCCACTGCCG 560  
QY 110 ataggctgacacaatatcttcttcaaaacttcaggaaagcaaatatagaaatganc 169  
Db 561 AAGAGCTACAAAAGACTTAGAAAAGCGCTTTTGACGATAATCAATATTCGGTATTGATC 620  
QY 170 atctactaggaaggaatcncagtnaaatcctncaggtttaagggtttcaaaannnagntt 229  
Db 621 ATTATCTTGGTAAAGAAATGTCGCAAAATATTGCTGCCCTTCGTTTGGTAAACCCCATCT 680  
QY 230 ttgagccaccttgagngnagcnnnnnnnganna 263  
Db 681 TTGATGCCGCTTGGACAAAGATTACATTAAAAA 714

RESULT 9  
US-08-445-464C-3  
Sequence 3, Application US/08445464C  
Patent No. 6090567  
GENERAL INFORMATION:  
APPLICANT: Jakobovits, Edward B.  
APPLICANT: Sillen, Joy L.  
APPLICANT: Levy, Mark J.  
APPLICANT: Goodman, Thomas C.  
APPLICANT: Becker, Martin  
APPLICANT: Ullman, Edwin F.  
APPLICANT: Caldwell, Robert M.  
APPLICANT: Bott, Richard R.  
APPLICANT: Barnett, Christopher C.  
TITLE OF INVENTION: Homogenous Immunoassays Using Mutant  
Glucose-6-Phosphate Dehydrogenases  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dade Behring Inc.  
STREET: 1717 Deerfield Road  
CITY: Deerfield  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60015  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445.464C  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/044,857  
FILING DATE: 08-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Ruzsala, Lois K.  
REGISTRATION NUMBER: 39,074  
REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (847) 267-5364  
TELEFAX: (847) 267-5376  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1571 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-445-464C-3

Query Match 14.1%; Score 37.2; DB 5; Length 1571;  
Best Local Similarity 44.9%; Pred. No. 0.007;  
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 50 agaagggatggaatcgcaataatttgaagccatttgcttggatgcactttcttccc 109  
Db 501 ATAGTGGTACACCGTTTGATGATTGAAAGCCCTTTGGTACATCATACGCCACTGCCG 560  
QY 110 ataggctgacacaatatcttcttcaaaacttcaggaaagcaaatatagaaatganc 169  
Db 561 AAGAGCTACAAAAGACTTAGAAAAGCGCTTTTGACGATAATCAATATTCGGTATTGATC 620  
QY 170 atctactaggaaggaatcncagtnaaatcctncaggtttaagggtttcaaaannnagntt 229  
Db 621 ATTATCTTGGTAAAGAAATGTCGCAAAATATTGCTGCCCTTCGTTTGGTAAACCCCATCT 680  
QY 230 ttgagccaccttgagngnagcnnnnnnnganna 263  
Db 681 TTGATGCCGCTTGGACAAAGATTACATTAAAAA 714

RESULT 10  
US-07-737-071A-1  
Sequence 1, Application US/07737071A  
Patent No. 529286  
GENERAL INFORMATION:  
APPLICANT: JARSCH, Michael  
APPLICANT: LANG, Gunter  
TITLE OF INVENTION: CLONING AND OVEREXPRESSION OF  
GLUCOSE-6-PHOSPHATE DEHYDROGENASE FROM LEUCONOSTOC  
DEXTRANICUS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Armstrong, Nikaido, Marmelstein Kubovcik &  
Murray  
STREET: 1725 K Street, N.W., Suite 1000  
CITY: Washington D.C.  
COUNTRY: United States of America  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/737,071A  
FILING DATE: 19910730  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 4024158.9  
FILING DATE: 30-JUL-1990  
ATTORNEY/AGENT INFORMATION:

```
; NAME: Murray, Robert B.
; REGISTRATION NUMBER: 22,890
; REFERENCE/DOCKET NUMBER: 911028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1696 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1580
; US-07-737-071A-1

Query Match          14.1%; Score 37.2; DB 1; Length 1696;
Best Local Similarity 44.9%; Pred. No. 0.0072;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 50 agaaggatggaatcgacataatatttgagaagccatttggttgcattgcacatttctccc 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 AGACTGGCTACAACTGTTGATGATTGAAAGCGCTTTTGGTACATCATACGCCACCGCAG 597

QY 110 ataggctgcacataatcttcttcaaaccttcagaaaagcaaatatagaaattganc 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 AAGAATTGCAAAAGTGATTGGGAAATGCAATTTGATGATGACCAACTGTCGTATTGACC 657

QY 170 atctactaggaagaatcncagtnaaatcctncaggttttaagggtttcaaaannagntt 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 ACTATCTTGGAAAGAAATGTTGACAAATATTGCAGCATTACGTTTGTGTAACCCAATCT 717

QY 230 ttgagccaccttngagngnacntnnnnnganna 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 TTGATCGCGCTTGGATAAGGACTATATCAAAAA 751

RESULT 11
US-08-022-096-1
; Sequence 1, Application US/08022096
; Patent No. 5308770
; GENERAL INFORMATION:
; APPLICANT: JARSCH, Michael
; APPLICANT: LANG, Gunter
; TITLE OF INVENTION: CLONING AND OVEREXPRESSION OF
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASE FROM LEUCONOSTOC
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/022,096
; FILING DATE: 19930225
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 40 24 158.0
; FILING DATE: 30-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/737,071
```

```
; FILING DATE: 30-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Chin, Monica F.
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-3005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1696 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1580
; US-08-022-096-1

Query Match          14.1%; Score 37.2; DB 1; Length 1696;
Best Local Similarity 44.9%; Pred. No. 0.0072;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 50 agaaggatggaatcgacataatatttgagaagccatttggttgcattgcacatttctccc 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 AGACTGGCTACAACTGTTGATGATTGAAAGCGCTTTTGGTACATCATACGCCACCGCAG 597

QY 110 ataggctgcacataatcttcttcaaaccttcagaaaagcaaatatagaaattganc 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 AAGAATTGCAAAAGTGATTGGGAAATGCAATTTGATGATGACCAACTGTCGTATTGACC 657

QY 170 atctactaggaagaatcncagtnaaatcctncaggttttaagggtttcaaaannagntt 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 ACTATCTTGGAAAGAAATGTTGACAAATATTGCAGCATTACGTTTGTGTAACCCAATCT 717

QY 230 ttgagccaccttngagngnacntnnnnnganna 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 TTGATCGCGCTTGGATAAGGACTATATCAAAAA 751

RESULT 12
US-08-445-463B-7
; Sequence 7, Application US/08445463B
; Patent No. 6033890
; GENERAL INFORMATION:
; APPLICANT: Jakobovits, Edward B.
; APPLICANT: Silen, Joy L.
; APPLICANT: Levy, Mark J.
; APPLICANT: Goodman, Thomas C.
; APPLICANT: Becker, Martin
; APPLICANT: Ullman, Edwin F.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Bott, Richard R.
; APPLICANT: Barnett, Christopher C.
; TITLE OF INVENTION: Homogenous Immunoassays Using Mutant
; TITLE OF INVENTION: Glucose-6-Phosphate Dehydrogenases
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dade Behring Inc.
; STREET: 1717 Deerfield Road
; CITY: Deerfield
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60015
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,463B
```

; FILING DATE: 22-MAY-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/044,857  
; FILING DATE: 08-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ruzsala, Lois K.  
; REGISTRATION NUMBER: 39,074  
; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (847) 267-5364  
; TELEFAX: (847) 267-5376  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1696 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-445-463B-7

Query Match 14.1%; Score 37.2; DB 5; Length 1696;  
Best Local Similarity 44.9%; Pred. No. 0.0072;  
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 50 agaagggatggaatcgcaataattttgagaagccatttgcttgatgcactttcttccc 109  
DB 538 AGACTGGCTACATCGTTTGATGATTGAAAGCCCTTTTGGTACATCATACGCCACCGCAG 597  
QY 110 atagggcgacacaataattctttcaaacctttcaggaagcaataatagaaattganc 169  
DB 598 AAGAATTGCAAGTGATTGGAAATGCATTGTGATGACCAACTGTCCTGATTGACC 657  
QY 170 atctactaggaagaatcncagtnaaatcctncaggttttaaggggtttcaaanngntt 229  
DB 658 ACTATCTTGGAAAGAAATGGTACAAATAATTGCGACATTACGTTTGGTAACCCAACT 717  
QY 230 ttgagccaccttngagngnacntnnnnnganna 263  
DB 718 TTGATCGCGCTTGGAAATGAGACTATATCAAAAA 751

RESULT 13  
US-08-445-464C-7  
; Sequence 7, Application US/08445464C  
; Patent No. 6090567  
; GENERAL INFORMATION:  
; APPLICANT: Jakobovits, Edward B.  
; APPLICANT: Silen, Joy L.  
; APPLICANT: Levy, Mark J.  
; APPLICANT: Goodman, Thomas C.  
; APPLICANT: Becker, Martin  
; APPLICANT: Ullman, Edwin F.  
; APPLICANT: Caldwell, Robert M.  
; APPLICANT: Bott, Richard R.  
; APPLICANT: Barnett, Christopher C.  
; TITLE OF INVENTION: Homogenous Immunoassays Using Mutant  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dade Behring Inc.  
; STREET: 1717 Deerfield Road  
; CITY: Deerfield  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60015  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,464C  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/044,857  
; FILING DATE: 08-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ruzsala, Lois K.  
; REGISTRATION NUMBER: 39,074  
; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (847) 267-5364  
; TELEFAX: (847) 267-5376  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1696 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-445-464C-7

Query Match 14.1%; Score 37.2; DB 5; Length 1696;  
Best Local Similarity 44.9%; Pred. No. 0.0072;  
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 50 agaagggatggaatcgcaataattttgagaagccatttgcttgatgcactttcttccc 109  
DB 538 AGACTGGCTACATCGTTTGATGATTGAAAGCCCTTTTGGTACATCATACGCCACCGCAG 597  
QY 110 atagggcgacacaataattctttcaaacctttcaggaagcaataatagaaattganc 169  
DB 598 AAGAATTGCAAGTGATTGGAAATGCATTGTGATGACCAACTGTCCTGATTGACC 657  
QY 170 atctactaggaagaatcncagtnaaatcctncaggttttaaggggtttcaaanngntt 229  
DB 658 ACTATCTTGGAAAGAAATGGTACAAATAATTGCGACATTACGTTTGGTAACCCAACT 717  
QY 230 ttgagccaccttngagngnacntnnnnnganna 263  
DB 718 TTGATCGCGCTTGGAAATGAGACTATATCAAAAA 751

RESULT 14  
US-08-933-821-16  
; Sequence 16, Application US/08933821  
; Patent No. 5972338  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; TITLE OF INVENTION: Tie Ligands  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,821  
; FILING DATE:  
; CLASSIFICATION: 536

	121	caatctctcttccaacttcaggaaaagcaaatatagaattgancatcttactagga	181
Qy	121	caatctctcttccaacttcaggaaaagcaaatatagaattgancatcttactagga	181
	589	CAATATTAACAACTTAACCAACGATAGTCAATTAAGCAANTAGAAATCGACTCAGA	648
Db	589	CAATATTAACAACTTAACCAACGATAGTCAATTAAGCAANTAGAAATCGACTCAGA	648
Qy	181	aggaatcncagtgnaaaatcctctncaag	205
Db	649	AGGACTAGTATTCAAGAACCCACAG	673

Search completed: November 4, 2000, 13:33:21  
Job time: 16637 sec

Query Match 12.5%; Score 33; DB 4; Length 2042;  
Best Local Similarity 60.0%; Pred. No. 0.17;  
Matches 51: Conservative 0; Mismatches 34; Indels 0; Gaps 0;

[illegible]

QY 181 aggaatcncagtnaaatcctncag 205  
||||| | | | | | | | | |  
Db 649 AGGACTAGTATTCAAGAAACCCACAG 673

RESULT 15

US-08-934-494-5  
: Sequence 5, Application US/08934494

Patent No. 6030831  
GENERAL INFORMATION:

: APPLICANT: Gurney, Austin

**TITLE OF INVENTION: Tie Ligands**

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 1 DNA way

CITY: South San Francisco

STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:  
SYSTEM TYPE: 2 5 disk 1 44 mb floppy disk

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatix (Genetech)

```

; SOFTWARE: winpatin (Genentech)
: CURRENT APPLICATION DATA:

```

```

;
; CURRENT APPLICATION DATA:
: APPLICATION NUMBER: IIS/08/934.494

```

APPLICATION NUMBER: 03/06/554,434  
FILING DATE:

FILE NO. : 100-441101-1000

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33

REFERENCE/DOCKET NUMBER: P10

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-3216

TELEFAX: 650/952-9881

```

; INFORMATION FOR SEQ ID NO: 5:
-----

```

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2042 base pairs
TYPE: nucleic acid

```

```

;
TYPE: Nucleic Acid
EXPANDEDNESS: single

```

; STRANDEDNESS: Single  
; TOPOLOGY: Linear

TOPLOGY: LINEAR  
; ITS-08-934-494-5

US-08-934-494-3

```
Query Match      12.5%; Score 33; DB 5; Length 2042;
Best Local Similarity 60.0%; Pred. No. 0.17;
Matches 51; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

Query Match 12.58; Score 33; DB 5; Length 2042;

Best Local Similarity 60.0%; Pred. No. 0.17;

Matches	51;	Conservative	0;	Mismatches	34;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:51:59 ; Search time 4352.3 Seconds  
(without alignments)  
373.614 Million cell updates/sec

Title: US-09-300-482-4  
Perfect score: 263

Sequence: 1 gaagcacttttgatgtgc.....agngnacntnnnnnganna 263

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estcom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estpl1.\*  
71: em\_estpl2.\*  
72: em\_estpl3.\*  
73: em\_estpl4.\*  
74: em\_estpl5.\*  
75: em\_estrol.\*  
76: em\_estrol2.\*  
77: em\_estrol3.\*  
78: em\_estrol4.\*  
79: em\_estrol5.\*  
80: em\_estrol6.\*  
81: em\_estrol7.\*  
82: em\_estrol8.\*  
83: em\_estrol9.\*  
84: em\_estrol10.\*  
85: em\_estrol11.\*  
86: em\_estrol12.\*  
87: em\_estrol13.\*  
88: gb\_gss1.\*  
89: gb\_gss2.\*  
90: gb\_gss3.\*  
91: gb\_gss4.\*  
92: em\_gss1.\*  
93: em\_gss2.\*  
94: em\_gss3.\*  
95: em\_gss4.\*  
96: gb\_gss5.\*  
97: gb\_gss6.\*  
98: gb\_gss7.\*  
99: gb\_gss8.\*  
100: gb\_gss9.\*  
101: em\_gss5.\*  
102: em\_gss6.\*  
103: em\_gss7.\*  
104: em\_gss8.\*  
105: em\_gss9.\*  
106: em\_gss10.\*  
107: em\_gss11.\*  
108: gb\_gss10.\*  
109: gb\_gss11.\*  
110: em\_gss12.\*  
111: gb\_gss12.\*  
112: gb\_gss13.\*  
113: gb\_gss14.\*  
114: gb\_gss15.\*  
115: gb\_gss16.\*  
116: gb\_gss17.\*

117: gb\_gss18:.\*  
118: gb\_gss19:.\*  
119: em\_gss13:.\*  
120: gb\_gss20:.\*  
121: gb\_gss21:.\*  
122: gb\_gss22:.\*  
123: gb\_gss23:.\*  
124: gb\_gss24:.\*  
125: em\_gss14:.\*  
126: em\_gss15:.\*  
127: em\_gss16:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	141.4	53.8	443	14	AL370545 MtBA38E01
2	50.6	19.2	295	21	AW335839 21995 MAR
3	48.8	18.6	260	36	BE526947 M66J03STM
4	48.2	18.3	624	19	AV561737 AV561737
5	48.2	18.3	624	19	AV561737 AV561737
6	46.6	17.7	382	13	AI894720 EST264163
7	46.6	17.7	382	13	AI894720 EST264163
8	46.6	17.7	382	13	AI894720 EST264163
9	45.2	17.2	661	23	AW686120 NF038D03N
10	45.2	17.2	661	23	AW686120 NF038D03N
11	43.2	16.4	360	22	AW479122 23661 MAR
12	43.2	16.4	360	22	AW479122 23661 MAR
13	42.6	16.2	385	24	AW784451 2622911.9
14	42.6	16.2	385	24	AW784451 2622911.9
15	42.6	16.2	385	24	AW784451 2622911.9
16	42.6	16.2	385	24	AW784451 2622911.9
17	41.8	15.9	695	33	BE548025 601072053
18	41.8	15.9	695	33	BE548025 601072053
19	40.2	15.3	615	5	BE515829 WHE0606.B1
20	40.2	15.3	615	5	BE515829 WHE0606.B1
21	38.8	14.8	881	34	BE216967 EST0510.T
22	37.2	14.1	712	34	BE262862 601147956
23	37.2	14.1	712	34	BE262862 601147956
24	37.2	14.1	712	34	BE262862 601147956
25	36.8	14.0	392	18	AL199390 Tetraodon
26	36.8	14.0	392	18	AL199390 Tetraodon
27	36.8	14.0	392	18	AL199390 Tetraodon
28	35.8	13.6	850	35	BE412598 MCG002.B1
29	35.8	13.6	850	35	BE412598 MCG002.B1
30	35.6	13.5	517	36	C25654 C25654 Dict
31	35.4	13.5	526	6	AA802117 GM03881.5
32	35.4	13.5	533	6	AA802117 GM03881.5
33	35.4	13.5	533	6	AA802117 GM03881.5
34	35.4	13.5	533	6	AA802117 GM03881.5
35	34.8	13.2	256	36	BE524751 M53H3STM
36	34.6	13.2	302	28	BB246584 BB246584
37	34.2	13.0	489	24	AW736245 EST33231
38	34.2	13.0	489	24	AW736245 EST33231
39	34.2	13.0	489	24	AW736245 EST33231
40	34.2	13.0	489	24	AW736245 EST33231
41	34.2	13.0	489	24	AW736245 EST33231
42	34.2	13.0	489	24	AW736245 EST33231
43	34.2	13.0	489	24	AW736245 EST33231
44	34.2	13.0	489	24	AW736245 EST33231
45	34.2	13.0	489	24	AW736245 EST33231

# ALIGNMENTS

RESULT 1  
AL370545

LOCUS DEFINITION AL370545 443 bp mRNA EST 03-AUG-2000  
MTBA38E01F1 MtBA Medicago truncatula cDNA clone M1830501-137, mRNA sequence.  
ACCESSION AL370545  
VERSION AL370545.1 GI:9670298  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
REFERENCE 1 (bases 1 to 443)  
AUTHORS Journal,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.  
TITLE Medicago truncatula ESTs from nitrogen-starved roots  
JOURNAL Unpublished (2000)  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html)  
FEATURES  
Location/Qualifiers  
1..443  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/db\_xref="taxon:3880"  
/clone="MTBA38E01"  
/clone\_lib="MtBA"  
/tissue\_type="root tips"  
/dev\_stage="harvested after 3 days of N-starvation"  
/note="Vector: pBluescript PSK; Site:1: EcoRI; Site:2: XhoI; Plants were grown in an aeroponic chamber for 14 days on nitrogen-rich medium followed by 3 days on N-free medium. RNA was extracted from root tips (1-3 cm). cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapXR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."  
BASE COUNT 131 a 96 c 94 g 122 t  
ORIGIN  
Query Match 53.8%; Score 141.4; DB 14; Length 443;  
Best Local Similarity 87.5%; Pred. No. 7.6e-31;  
Matches 154; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
Qy 1 gaagcatttttgatgttgatgtcttgcgaagcagtcctcagaccagagatgag 60  
Db 200 GAAGCATTGTTGGACGTGCTTTCATGTTGCAAGCAGTCTCAGACCCAAAGGATGC 259  
Qy 61 aatgcataatttgagaagcatttgcttgatgcacttttctccatagctgaca 120  
Db 260 ATTCGATATAATGAGAGCCATTGGATTGATGCTCTCTCTCCCAAGGCTGACA 319  
Qy 121 caatatcttttcaactttcagaaagcaaatatataagaatgancatctact 176  
Db 320 CAATATCTCTCTCGAAGTTTGAGAAAGCACTATATAGGTATTCATACCTTCT 375  
RESULT 2  
AW335839 295 bp mRNA EST 10-JUL-2000  
LOCUS DEFINITION 21995 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AW335839







BASE COUNT 102 a 72 c 87 g 121 t  
ORIGIN

Query Match 17.7%; Score 46.6; DB 13; Length 382;  
Best Local Similarity 47.8%; Pred. No. 0.0023;  
Matches 100; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 55 ggatgaatcgcataataatttgagagccattggcttgatgcacttttcccatagg 114  
Db 113 GGTGGACAGCAATCGTTGTGAGAGCCCTTCGGCAAGGATTTAGCTTCATCTGAACAA 172

QY 115 ctgacacaatatctcttcaaaccttcaggaaaagcaaatatataagaattgancatcta 174  
Db 173 CTAAGTCCAGATCGGAGAACTATTTCACCAACCCCAAAATTTATCCATTTGACCATAT 232

QY 175 ctgagaagaatcncagtnaaatctcncaggttttaagggtttcaaanngntttgag 234  
Db 233 TTGGGAAGGAGTGGTTGCAGAAATTTGTGTGCTGCGTTTTCGCAAAATCGCTTCTTTTG 292

QY 235 ccaccttngagngnnaacntnnnnnganna 263  
Db 293 CGCTCTGGAATCGTGACAACTTGATAA 321

## RESULT 7

AW686120 649 bp mRNA EST 29-JUN-1999  
LOCUS EST241911 tomato shoot, Cornell Lycopersicon esculentum cDNA clone  
DEFINITION CLEB1G14 similar to glucose-6-phosphate dehydrogenase, mRNA  
sequence.

ACCESSION AI491202  
VERSION AI491202.1 GI:4386512  
KEYWORDS EST.  
SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids  
I; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 649)

AUTHORS van der Hoeven, R.S., Matern, A.L., Vision, T., Holt, I.E., Liang, F.,  
Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L.,  
Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Giovannoni, J.J.  
and Tanksley, S.D.

TITLE Generation of ESTs from tomato shoot meristem

JOURNAL Unpublished (1999)

COMMENT Contact: David Frisch

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 4366

Fax: 864 656 4293

Email: dfrisch@CLEMSON.EDU.

FEATURES Location/Qualifiers

source 1..649

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="CLEB1G14"

/clone\_lib="tomato shoot, Cornell"

/tissue\_type="shoot meristem"

/dev\_stage="8 week old plants"

/lab\_host="XLOLR"

/note="Vector: pBK CMV; Site 1: EcoRI; Site 2: XhoI; CLEB  
- Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA  
library made from tomato vegetative shoots including  
meristems and small expanding leaves."

BASE COUNT 182 a 117 c 154 g 196 t

ORIGIN

Query Match 17.7%; Score 46.6; DB 11; Length 649;  
Best Local Similarity 47.8%; Pred. No. 0.0026;

Matches 100; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 55 ggatgaatcgcataataatttgagagccattggcttgatgcacttttcccatagg 114  
Db 360 GGTGGACAGCAATCGTTGTGAGAGCCCTTCGGCAAGGATTTAGCTTCATCTGAACAA 419

QY 115 ctgacacaatatctcttcaaaccttcaggaaaagcaaatatataagaattgancatcta 174  
Db 420 CTAAGTCCAGATCGGAGAACTATTTCACCAACCCCAAAATTTATCCATTTGACCATAT 479

QY 175 ctgagaagaatcncagtnaaatctcncaggttttaagggtttcaaanngntttgag 234  
Db 480 TTGGGAAGGAGTGGTTGCAGAAATTTGTGTGCTGCGTTTTCGCAAAATCGCTTCTTTTG 539

QY 235 ccaccttngagngnnaacntnnnnnganna 263  
Db 540 CGCTCTGGAATCGTGACAACTTGATAA 568

## RESULT 8

AW686120 661 bp mRNA EST 15-JUN-2000  
LOCUS NF038D09NR1F1000 Nodulated root Medicago truncatula cDNA clone  
DEFINITION NF038D09NR 5', mRNA sequence.

ACCESSION AW686120  
VERSION AW686120.1 GI:7560856  
KEYWORDS EST.  
SOURCE barrel medic.

ORGANISM Medicago truncatula

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euasterids I;  
Fabales; Fabaceae; Papilionoideae; Medicago.  
1 (bases 1 to 661)

AUTHORS Watson, B.S., Shin, H.-S., Lopez-Meyer, M., Scott, A.D., Harris, A.R.,  
Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W., May,  
G.D. and Paiva, N.L.

TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula nodulated root library  
JOURNAL Unpublished (2000)

COMMENT Contact: Paiva NL

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7317

Fax: 580 221 7380

Email: nlpaiva@noble.org

Insert Length: 661 Std Error: 0.00

Plate: 038 row: D column: 09

Seq primer: TCACACAGGAACACGCTATGAC.

FEATURES

source

1..661

/organism="Medicago truncatula"

/db\_xref="taxon:3880"

/clone="NF038D09NR"

/clone\_lib="Nodulated root"

/tissue\_type="root"

/dev\_stage="pooled developmental"

/note="Vector: Lambda Zap; Four-week old Rhizobium  
melliloti-inoculated Medicago truncatula roots, containing  
a mixture of young and old roots and nodules."

BASE COUNT 194 a 116 c 152 g 199 t

ORIGIN

Query Match 17.7%; Score 46.6; DB 23; Length 661;  
Best Local Similarity 47.8%; Pred. No. 0.0026;  
Matches 100; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 55 ggatgaatcgcataataatttgagagccattggcttgatgcacttttcccatagg 114  
Db 37 GGATGACACGCGTTGTGTTGAGAAACCCCTTTGGTAGGATCTAGATCTGCGAAGAA 96

QY 115 ctgacacaatatctcttcaaaccttcaggaaaagcaaatatataagaattgancatcta 174

```

Db 97 CTCAGTACTCAGATTGGAGAGTATTATTGGAAGAACCAAAATTTATCGTATTGATCACAT 156
Qy 175 ctagaaggaatcncagtaaaatcncaggtttaagggttcaaanngnntttgag 234
Db 157 TTGGGAAGGAAGTGTGCAAAACATGTTAGTACTTCGTTTTCGCAAAATCGGTTCTCTTG 216
Qy 235 ccaccttngagngnnaacntnnnnnganna 263
Db 217 CCTCTTTGGAACCAACCAACCATTTGACAA 245

RESULT 9
LOCUS AW180861 735 bp mRNA EST 17-NOV-1999
DEFINITION MGA1030f Mga Library Mycosphaerella graminicola cDNA clone Mga1030
5' similar to GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE, mRNA sequence.
ACCESSION AW180861
VERSION AW180861.1 GI:6448056
KEYWORDS EST.
SOURCE Mycosphaerella graminicola.
ORGANISM Mycosphaerella graminicola
Eukaryota; Fungi; Ascomycota; Ascomycota incertae sedis;
Mycosphaerellaceae; Mycosphaerella.
REFERENCE 1 (bases 1 to 735)
AUTHORS Keon, J.P.R., Bailey, A.M. and Hargreaves, J.A.
TITLE A group of expressed cDNA sequences from the wheat fungal leaf
blotch pathogen, Mycosphaerella graminicola
JOURNAL Unpublished (1999)
COMMENT Contact: Hargreaves JA
Cell Biology Department
IACR-Long Ashton Research Station
Long Ashton, Bristol, BS41 9AF, UK
Tel: +44(0)1275 392181
Fax: +44(0)1275 394281
Email: john.hargreaves@bbsrc.ac.uk
Insert Length: 1100 Std Error: 0.00
Seq Primer: M13 reverse.

FEATURES
source
1..735
Location/Qualifiers
/organism="Mycosphaerella graminicola"
/strain="Strit"
/db_xref="taxon:54734"
/clone="Mga1030"
/clone_lib="Mga Library"
/note="Vector: pSPORT1; Library constructed from cultures
utilizing ammonium ions as a source of nitrogen"
BASE COUNT 181 a 212 c 189 g 153 t
ORIGIN

Query Match 17.2%; Score 45.2; DB 20; Length 735;
Best Local Similarity 47.2%; Pred. No. 0.0068;
Matches 101; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 50 aagaaggatggaatcgataattttgagaagccatttggttgatgcattttctccc 109
Db 10 AGAACGGGCAATTCGCCGATCATCATCGAGAGAGCCCTTCGGCAAGAGACTCCAGTCGTCG 69
Qy 110 ataggctgacacaatatctcttcttcaaaccttcaggaaagcaaatatagattganc 169
Db 70 GCGGCTCTCGACAAAGCTCTCCGCCCAACTTGAAGGAGGAGAGATCTTCGCGATCGATC 129
Qy 170 atctactagaaggaatcncagtaaaatcncaggtttaagggttcaaanngnntt 229
Db 130 ACTATCTCGGAAGGAGATGGTCAAGAACATTTCTCATTTCTCGTTTCGGAACCAATTC 189
Qy 230 ttgagccacttngagngnnaacntnnnnnganna 263
Db 190 TTGGCCTACTTGGATCCGACCAACCATCGATAA 223

RESULT 10

```

```

AU095224 465 bp mRNA EST 30-JUN-2000
LOCUS AU095224 Rice cDNA from immature leaf including apical meristem
DEFINITION (under short day condition) Oryza sativa cDNA clone E60680, mRNA
sequence.
ACCESSION AU095224
VERSION AU095224.1 GI:8857906
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 465)
AUTHORS Sasaki, T. and Yamamoto, K.
TITLE Rice cDNA from immature leaf including apical meristem (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.
E60680_1A. Location/Qualifiers
1..465
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E60680"
/clone_lib="Rice cDNA from immature leaf including apical
meristem (under short day condition)"
/dev_stage="immature leaf including apical meristem (under
short day condition)"
BASE COUNT 130 a 83 c 112 g 135 t 5 others
ORIGIN

Query Match 17.1%; Score 45; DB 15; Length 465;
Best Local Similarity 47.4%; Pred. No. 0.0069;
Matches 99; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 55 ggaatgaatcgataattttgagaagccatttggttgatgcattttctcccattagg 114
Db 35 GGATGGAGTACAGATGCTATTGTTGAGAGCCCATTTGGAAGGAGACTTGGACTCTGCAAGAA 94
Qy 115 ctgacacaatatcttcttcaaaccttcaggaaagcaaatatagattgancatcta 174
Db 95 TTAAGTGCCCAACTTGGGGAGCTATTGACGAAACCAACTCTACAGAAATTGATCATAC 154
Qy 175 ctagaaggaatcncagtaaaatcncaggtttaagggttcaaanngnntttgag 234
Db 155 TTGGGAAGAAGATTGGTTCACAAACCTGCTGCTGCTTTCGTTTGCACACCCCTGTTCTTG 214
Qy 235 ccaccttngagngnnaacntnnnnnganna 263
Db 215 CCTCTTTGGAACCGTGACAATATTGATAA 243

RESULT 11
LOCUS AW479122 360 bp mRNA EST 07-JUL-2000
DEFINITION 23661 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW479122
VERSION AW479122.1 GI:7049228
KEYWORDS EST.
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

```



Sat Nov 4 18:11:40 2000

Seq primer: -40M13RevUniv  
High quality sequence stop: 385.

FEATURES  
source

1. .385  
Location/Qualifiers  
/organism="Canis familiaris"  
/db\_xref="taxon:9615"  
/clone="2b62911"  
/clone\_lib="Canis cDNAs from mdck cells"  
/note="Vector: Lambda Zap II; The library was provided by  
Greg Hannon (Cold Spring Harbor Laboratory). This  
library is oligo(dT) primed using Stratagene zap cDNA  
synthesis kit. It was made from exponentially growing mdck  
cells. Please contact Greg Hannon (hannon@esl.org) with  
any library related inquiries."  
85 a 115 c 111 g 74 t

BASE COUNT  
ORIGIN

Query Match 16.2%; Score 42.6; DB 24; Length 385;  
Best Local Similarity 48.5%; Pred. No. 0.033;  
Matches 99; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 45 gaccacgaaggatggaatcgataatatttgagaagccatttgcttgatgcactttc 104  
DB 39 GAGCCACACAGGCTGGAAACCGTGTCTCGTGGAGAGGCCCTTCGGAGGAGCCTGCAGAG 98  
QY 105 ttcccataggctgacacataatcttcttcaaacattttcaggaagcaaaatataagaat 164  
DB 99 CTCGACAGGCTGTCCACACCATCTCGTCTTGTTCGCGAGGAGCAGATCTACCGCAT 158  
QY 165 tgancatctactaggaagaatcncagtnaaatcctncaggtttaagggtttcaaan 224  
DB 159 GCACCACTACTCTGGCAAGAGATGGTCCAGAACCTCATGTGTGCTGAGTTTGCCACAG 218  
QY 225 agnttttgagccacctngagng 248  
DB 219 GATCTTCGGCCCATCTGGAACCG 242

RESULT 14  
BE264631  
LOCUS

BE264631 565 bp mRNA EST 13-JUL-2000  
601192423F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3536453 5',  
mRNA sequence.

ACCESSION  
VERSION

BE264631  
BE264631.1 GI:9138189  
EST.

KEYWORDS  
SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 565)  
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

FEATURES  
source

1. .565  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3536453"  
/clone\_lib="NIH\_MGC\_7"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT  
ORIGIN

Query Match 16.2%; Score 42.6; DB 34; Length 565;  
Best Local Similarity 48.5%; Pred. No. 0.036;  
Matches 99; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 45 gaccacgaaggatggaatcgataatatttgagaagccatttgcttgatgcactttc 104  
DB 303 GAGCCAGATAGCTGGAAACCGCATCTCGTGGAGAGGCCCTTCGGAGGAGCCTGCAGAG 362  
QY 105 ttcccataggctgacacataatcttcttcaaacattttcaggaagcaaaatataagaat 164  
DB 363 CTCTGACCGGCTGTCCACACCATCTCTCTCCCTGTTCGCTGAGGAGCAGATCTACCGCAT 422  
QY 165 tgancatctactaggaagaatcncagtnaaatcctncaggtttaagggtttcaaan 224  
DB 423 CGACCACTACTCTGGCAAGAGATGGTCCAGAACCTCATGTGTGCTGAGATTTGCCACAG 482  
QY 225 agnttttgagccacctngagng 248  
DB 483 GATCTTCGGCCCATCTGGAACCG 506

## RESULT 15

BE548025  
LOCUS

BE548025 782 bp mRNA EST 09-AUG-2000  
601072053F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:3458028 5',  
mRNA sequence.

ACCESSION  
VERSION

BE548025  
BE548025.1 GI:9776670  
EST.

KEYWORDS  
SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 782)  
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

REFERENCE  
AUTHORS

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM8448 row: j column: 13  
High quality sequence stop: 738.  
Location/Qualifiers  
1. .782  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3458028"  
/clone\_lib="NIH\_MGC\_12"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: placenta; Vector: pCMV-SPORT6; Site:1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.4 kb. Library prepared by Life  
Technologies."

FEATURES  
source

1. .782  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3458028"  
/clone\_lib="NIH\_MGC\_12"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: placenta; Vector: pCMV-SPORT6; Site:1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.4 kb. Library prepared by Life  
Technologies."

BASE COUNT  
ORIGIN

172 a 242 c 217 g 150 t 1 others

	Query Match	16.2%	Score 42.6;	DB 36;	Length 782;
	Best Local Similarity	50.8%;	Pred. No. 0.039;		
	Matches	94;	Conservative	0;	Mismatches 91; Indels 0; Gaps 0;
QY	45	gaccagaagggatgaataatcgcaataattttgagaagccatttgggttttgatgcacttc	104		
Db	504	GAGCCAGATAGGCTGGAAACGCCATCATCGTGGAGAAGCCCTTCGGGAGSGACCCTGCAGAG	563		
QY	105	tccccataggctgacacaatatcttttccaaccttccgaaaaaacaaatatagaat	164		
Db	564	CTGTGACCGGCTGTCCAACCAATCTCCTCCCGTTCCGTGAGGACCAAGATCTAACGCAT	623		
QY	165	tgancattctactagaaggaatacncagtnaaatacctncaagttttaagggtttcaaann	224		
Db	624	CGANCACTACCTGGGCAAGGAGATGTTGTGAGAACCCTTCATGGTGCTGAGATTTCGAACAGG	683		
QY	225	agnnt	229		
Db	684	ATCTT	688		

Search completed: November 4, 2000, 11:52:01  
Job time: 18238 sec

100



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 08:49:11 ; Search time 4075.18 Seconds  
(without alignments)  
249.718 Million cell updates/sec

Title: US-09-300-482-1

Perfect score: 233

Sequence: 1 gttttgcagttagtagaat.....ttggacaagggtattgttg 233

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_bal.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pl1.\*

8: gb\_pl2.\*

9: gb\_pr1.\*

10: gb\_pr2.\*

11: gb\_pr3.\*

12: em\_fun.\*

13: em\_hum1.\*

14: em\_hum2.\*

15: em\_in.\*

16: em\_om.\*

17: em\_or.\*

18: em\_ov.\*

19: em\_pat.\*

20: em\_ph.\*

21: em\_pl.\*

22: em\_ro.\*

23: em\_sts.\*

24: em\_sy.\*

25: em\_un.\*

26: em\_vi.\*

27: gb\_htg1.\*

28: gb\_htg2.\*

29: gb\_in1.\*

30: gb\_in2.\*

31: em\_bal.\*

32: em\_ba2.\*

33: em\_hum3.\*

34: em\_hum4.\*

35: gb\_pr4.\*

36: gb\_htg3.\*

37: gb\_htg4.\*

38: gb\_htg5.\*

39: gb\_htg6.\*

40: gb\_htg7.\*

41: em\_htg1.\*

42: em\_htg2.\*

43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_v11.\*  
94: gb\_v12.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	131.6	56.5	1904	7	AB029456	AB029456 Triticum
2	131.6	56.5	1924	7	AB029455	AB029455 Triticum
3	131.6	56.5	1956	7	AB029454	AB029454 Triticum
4	106.6	45.8	1785	45	MSU18238	U18238 Medicago sa
5	101.8	43.7	1689	45	STG6PDH	X74421 S.tuberosum
6	100.8	43.3	1957	7	AF012863	AF012863 Petroselin
7	97	41.6	1821	45	NTTCG9	AJ001770 Nicotiana
8	96	41.2	1848	7	AF012862	AF012862 Petroselin
9	95.4	40.9	1852	45	NTTCG6	AJ001769 Nicotiana
10	95.4	40.9	1862	8	ATH010970	AJ010970 Arabidops
11	93.8	40.3	1690	8	ATH010971	AJ010971 Arabidops
12	78	33.5	63604	8	AP000381	AP000381 Arabidops

us-09-300-482-1.rqe

```

/cultivar="Chinese spring"
/isolate="root"
/db_xref="taxon:4565"
/clone="Tagpd3"
/dev_stage="seedling"
/notes="hexaploid"
153..1682
/gene="g6pdh"
153..1682
/gene="g6pdh"
/codon_start=1
/product="glucose-6-phosphate dehydrogenase"
/protein_id="BAA97664.1"
/db_xref="GI:8918506"
/translation="MAGTOSTASSROSSPNSLAKLELPLEKGCLTIIVVLGASGLAK
KKTFFALYHLEQGGVTHVGTARTNLSDDGRLGRIRAYLKGASEHVSSEELQML
IKVSGSYHSGEGEFLKNEISDYEMNNSSGSRRLFLALPPSVYPVCKMRTFYCM
SPNSRAGTWRIVEKPPFGLDSAEELSQGLPEEDQLYRIDHYLGKELVQNLLVL
RFRANLEPLAWNEDNDNIQIVRFDFDGRGGYFDXYGIIIRGIIONHLLQVFCFLVA
MERPVSLKPHIRDEKVKVLQSVNPIKDEEVLGQYQGYKEOPTVPDSDNSTPTFASIV
LRVHNERKEGVPPIKAGLAKNSKAEIRVQFQDPEDGDFCKCKQGRNEFVIRLQPSF
AMTKMLTVKKGPLEMATESELDLGMRYQDIP1PEAYERLIDTIRGQQQHFFVARD
ELKAAWQIFLPLHLDIAGKLKAVSYKPSGRGPFKEADELSEKGVQMOTHYTWIIPPTL
A"

```

Query Match 56.5%; Score 131.6; DB 7; Length 1904;  
Best Local Similarity 76.3%; Pred. No. 6.3e-29;  
Matches 177. Conservative 0; Mismatches 49; Indels 6; Gaps 1;

Qy 2 ttttgcagttagtagaataattgttagtggtcctctatgataggggtggaagatttgagtta 61  
nb 447 ttttgcagttagtagaataattgttagtggtcctctatgataggggtggaagatttgagtta 506  
nb 447 ttttgcagttagtagaataattgttagtggtcctctatgataggggtggaagatttgagtta 506

507 CTGAACAAGGAATATCAGATTATGAGATGCTCA-----AACAACTCAGAAGCTCCGT 560

122 agattattttattgcatggtcccatcagtcacccatcagtcagtcgagatgaaga 181

561 AGGCTCTTTTATTGGCATTGCCCTCCATCTGTCTACCCTTCAGTGTGCAAAATGATCCGA 620

182 tcatattgcatgattccatcttcacacacgcggttggacaagggttatgttg 233

621 ACATATTGCATGAGTCCAACTTCTCGCGCTGGATGGACTAGAGTAATTTGTTG 672

RESULT	2	
AB029455		
LOCUS	1924 bp	mrna
DEFINITION	Triticum aestivum g6pdh mRNA for glucose-6-phosphate dehydrogenase, complete cds, clone:tagpd2.	
AB029455		
ACCESSION	AB029455.1	GI:8918503
VERSION	glucose-6-phosphate dehydrogenase.	
KEYWORDS	Triticum aestivum (cultivar:Chinese spring, isolate:root) seedling	
SOURCE	cDNA to mRNA, clone:tagpd2.	
ORGANISM	Triticum aestivum	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.	
AUTHORS	1 (sites)	
TITLE	Nemoto,Y., Kawakami,N. and Sasakuma,T.	
JOURNAL	Isolation of novel early salt-responding genes from wheat (Triticum aestivum L.) by differential display	
REFERENCE	Theor. Appl. Genet. 98, 673-678 (1999)	
AUTHORS	2 (sites)	
TITLE	Nemoto,Y.	
JOURNAL	Molecular characterization of glucose-6-phosphate dehydrogenase (G6PDH) from wheat (Triticum aestivum L.): gene expression in response to salt stress	
JOURNAL	Unpublished (1999)	





Db	500	TTGGATAAGGAATATCTGAGCATGAAATATCAAGAAACAGTATTGAAGGATCGTCTCCG	559
Qy	122	agattattttattggcatgtccctccatcagtcctaccatcagtcagtcgagatgataaga	181
Db	560	AGCACTATTTTACCTTCGACTTCCTCTTCAGTATATCTCTCTGTCGAAATGATTAA	619
Qy	182	tcatattgcatgagtcacattcttcacacacccggttggacaagggttatgttg	233
Db	620	ACTTACTGATGAATAAATCTGATCTTGATGGATGACTCGTATTGTGTGTTG	671

RESULT	7	
NTTCG9		
LOCUS	1821 bp	PLN 17-AUG-1999
DEFINITION	Nicotiana tabacum mRNA for cytosolic glucose-6-phosphate dehydrogenase TCG9.	
ACCESSION	AJ001170	
VERSION		
KEYWORDS	glucose-6-phosphate dehydrogenase.	
SOURCE	common tobacco.	
ORGANISM	Nicotiana tabacum	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;	
Nicotiana		
1 (bases 1 to 1821)		
Wendt,U.K., Hauschild,R., Lange,C., Pietersma,M., Wenderoth,I. and von Schaeuwen,A.		
Evidence for functional convergence of redox regulation in G6PDH isoforms of cyanobacteria and higher plants		
Plant Mol. Biol. 40 (3), 487-494 (1999)		
99364543		
2 (bases 1 to 1821)		
von Schaeuwen,A.		
Direct Submission		
Submitted (18-SEP-1997)		
University Osnabrueck, Barbarastr. 11, D-49069 Osnabrueck, GERMANY		
Location/Qualifiers		
1 - 1821		
FEATURES		
source		

```

/cultivar="Samsun NN"
/db_xref="taxon:4097"
/cell_type="mesophyll"
/tissue_type="young sink leaves"
88..1623
/notes="TCG9"
/codon_start=1
/product="glucose-6-phosphate dehydrogenase"
/protein_id="CAA04993.1"
/db_xref="GI:3021510"
/db_xref="SPTREMBL:O65855"
/translation="MAASWCIEKGRSIRLDSFRDNINIPETGCLSIIVLGAAGDGLAK
KTFPLFNLYRGQFQSLQSNVHFYGARTKISDDDLRGRNGYLSQGENEEVSEFLQ
LKIYVSGSDSGEGFLDQKAATHEIAKNTSGSRRLFYFALPPSVYPSVGRMKIN
YCMKNSDGGWTRIVPEKFKDQASALQSQIGELFDEFQIYRIDHYLKGELVNL
LVLRANRFELPLMNRDINDINIQVIRDFDETCRGVGYFQYGIIRDIIONQLLVLC
LVAMEKPSQKPEHRLDRDEKVLQVMSLPIKDEEYVGYQYEDKIDPTVPDNTNFA
TWVLRTHNRWGVPEFKAKGAKNSKAEIRVQFQKVPDGDIPRCKQKGRNVRILQ
PSEAMYMKLTVKKPKGLEMTVQSELDLSYRGORYGVVPEAYERLIDLTIRGQQQHFV
RDELMKAAWEITPPLHRIDDDGVKPIPKPGSGRPAEALQNVNGYVQVTHGYICIP
PL"
515 a 340 c 434 g 532 t

```

[illegible]





QY	1	gtttttgcagtttagtaaatattgttagtgcctcctcatgatagggtggaagatttgagctt	60
Db	375	gtyttctgcagctgatttaagattgtgagtgggcccttattgatctgagagaggggttccagag	434
QY	61	attgaataggagcaattctctgagtatgagacttcagaaacaatgactcgggaagctaccg	120
Db	435	atttagacaaggcaattttcagacacgaaatcctccaaaatagtactgaagggtcttcttag	494
QY	121	cagattattttattgtgcattgcctccatcagcttaccatcagctatcgagatcgatgataag	180
Db	495	aagactgtttttattctgacattccaccgctctgtttattccttctgtatgacagatgatcaa	554
QY	181	atcatattgcattgagtcattcttccacacaccgggttgacaagggttattgttg	233
Db	555	gacgtgctgcattgaacaaatctgatctgttggtgagtgacacggattgtttgtgg	607
RESULT	12		
AP000381/c			
LOCUS	AP000381	63604 bp	DNA
DEFINITION	Arabidopsis thaliana genomic DNA,	chromosome 3,	TAC clone:K17E12.
ACCESSION	AP000381		
VERSION	AP000381.1	GI:5672506	
KEYWORDS	Arabidopsis thaliana (strain:Columbia) DNA,	clone_lib:Mitsui	TAC
SOURCE	clone:K17E12.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;		
	Brassicales; Brassicaceae; Arabidopsi		
REFERENCE	1 (sites)		
AUTHORS	Nakamura,Y.		
TITLE	Structural analysis of Arabidopsis thaliana chromosome 3. II.		
	Sequence features of the regions of 4,251,695 bp covered by ninety		
	P1, TAC and BAC clones		
JOURNAL	DNA Res 7, 217-221 (2000)		
REFERENCE	2 (bases 1 to 63604)		
AUTHORS	Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-JUL-1999) to the DDBJ/EMBL/GenBank databases.		
	Yasukazu Nakamura, Kazusa DNA Research Institute, Department of		
	Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan		
	(E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935,		
	Fax:81-438-52-3934)		
COMMENT	Address for correspondence: kaos@kazusa.or.jp		
	For the latest information on annotation of this clone, please see		
	http://www.kazusa.or.jp/kaos/cgi-bin/sqd.graph.cgi?c=K17E12		
	Genes with similarity to proteins in the databases are described in		
	'product' or 'note' qualifiers. Genes that have no significant		
	protein similarity are described as 'unknown protein'.		
	The software programs used to predict genes include: Grail		
	(Informatics Group, Oak Ridge National Laboratory,		
	http://compbio.ornl.gov/grail-1.3/),		
	GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),		
	NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of		
	Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and		
	SplicePredictor (Volker Brendel, Stanford University,		
	http://gremlin2.zool.lastate.edu/cgi-bin/sp.cgi).		
	Genes encoding tRNAs are predicted by tRNAscan-SE		
	(Sean Eddy, Washington University School of Medicine, St. Louis,		
	http://genome.wustl.edu/eddy/tRNAscan-SE/).		
	This sequence may not be the entire insert of this clone. It may be		
	shorter because we remove overlaps between neighboring submissions.		
	The 5' clone is MYF5 and the 3' clone is K1G2.		
FEATURES	Location/Qualifiers		
source	1. 63604		
	/organism="Arabidopsis thaliana"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="3"		
	/clone="K17E12"		
	/clone_lib="Mitsui TAC"		
CDS	complement(join(3373..3519,3594..3676),3763..3835,		









```
repeat_region /note="L1 element fragment"
35550..35630
repeat_region /note="L1 element fragment"
36315..36395
repeat_region /note="L1 element fragment"
36510..36657
repeat_region /note="L1 element fragment"
36700..36819
repeat_region /note="MER42A element fragment"
36705..36785
repeat_region /note="MER42C element fragment"
36841..37025
repeat_region /note="L1 element fragment"
37026..37184
repeat_region /partial
/note="Alu repeat: matches 1..174 of consensus"
BASE COUNT 12421 a 7079 c 6961 g 10721 t 2 others
ORIGIN

Query Match 16.7%; Score 39; DB 82; Length 37184;
Best Local Similarity 50.3%; Pred. No. 0.28;
Matches 96; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1 gtttttcagtttagaataatgttagtggtcctctatgataggtggaaggatttgatt 60
DB 8340 GCTTCTGCATTTGTAGATCATATATATCAACTTATTAATAATACAGGTGAGAGTCTCAA 8281

QY 61 attgaatgaggcaatctctgagtagatgagacttcagaaaaacaatgactcgggaagctaccg 120
DB 8280 ATAGCTAGCTGACATCTCAGTGGACTAGTAGTATCAACATTTGCTAGCTCTGTAGCATGGA 8221

QY 121 cagattattttatttgccatgcctccatcagtcacccatcagtagatgcgagatgataag 180
DB 8220 CAGATTATTTCAGTTATCTGTCTCTGTTTATTTCCTGTACATGGAGAGATAATATTAG 8161

QY 181 atcatatigca 191
DB 8160 TGCCTATCTCA 8150
```

Search completed: November 4, 2000, 13:30:03  
Job time: 16852 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 09:11:13 ; Search time 320.8 Seconds  
(without alignments)  
272.847 Million cell updates/sec

Title: US-09-300-482-1

Perfect score: 233

Sequence: 1 gttttgcagtttagtaaat.....ttggacaagggtattgttg 233

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31.6	13.6	278	21 A01024	Human colon cancer
2	31.2	13.4	4418	18 W74765	Staphylococcus aur
3	29.8	12.8	5280	18 W74855	Staphylococcus aur
C 4	29.8	12.8	5631	20 X85506	DNA encoding. Sac
C 5	29.8	12.8	7655	16 Q86896	S. cerevisiae fks1
C 6	29.2	12.5	5342	21 D00374	Rat smooth muscle-
C 7	28.8	12.4	5354	20 X13263	Enterococcus faeca
C 8	28.8	12.4	18613	18 W74423	Staphylococcus aur
C 9	28.6	12.3	7313	19 V29268	Nucleotide sequenc
C 10	28.4	12.2	3508	16 Q86693	Candida tropicalis
C 11	28.4	12.2	5333	18 W74481	Staphylococcus aur
C 12	28.4	12.2	17710	19 V31256	E. coli J96 pathog

C 13	28.2	12.1	11887	19 V52279	Streptococcus pneu
C 14	28.2	12.1	12565	19 V62392	Human interleukin-
C 15	28.2	12.1	12565	20 X75925	Human interleukin
C 16	28	12.0	337	14 Q59415	Human brain Expre
C 17	28	12.0	3417	17 T39622	Human DNA ligase I
C 18	28	12.0	3417	20 X82491	Human DNA ligase I
C 19	27.8	11.9	3050	20 X29768	L.lactis HsdR gene
C 20	27.6	11.8	32768	19 V52204	Streptococcus pneu
C 21	27.4	11.8	527	19 V31323	E. coli J96 pathog
C 22	27.4	11.8	1172	19 V53308	DNA encoding a Sta
C 23	27.4	11.8	1513	21 A26678	Candida albicans p
C 24	27.4	11.8	2442	20 Z23609	Human labyrinthin
C 25	27.4	11.8	5892	18 T79680	BRCA2 cancer suscep
C 26	27.4	11.8	7240	18 T79680	BRCA2 cancer suscep
C 27	27.4	11.8	8703	19 V52310	Partial BRCA2 cance
C 28	27.4	11.8	10485	20 X30256	Streptococcus pneu
C 29	27.4	11.8	10485	20 X30257	Human BRCA2 (Omi2)
C 30	27.4	11.8	10485	20 X30258	Human BRCA2 (Omi3)
C 31	27.4	11.8	10485	20 X30259	Human BRCA2 (Omi4)
C 32	27.4	11.8	10485	20 X30255	Human BRCA2 (Omi5)
C 33	27.4	11.8	11283	18 T84841	Human BRCA2 (Omi1)
C 34	27.4	11.8	11283	21 T87996	Human breast and o
C 35	27.4	11.8	11385	18 T69707	BRCA2 gene sequenc
C 36	27.2	11.7	439	21 A32071	Human breast cance
C 37	27.2	11.7	860	18 X30792	Plant microsatelli
C 38	27.2	11.7	1153	19 T98660	Streptococcus pneu
C 39	27.2	11.7	3764	21 Z44192	DNA encoding a S.
C 40	27	11.6	10240	19 V39007	Murine cerebral ne
C 41	26.8	11.5	690	21 A01923	Porcine INF-alpha
C 42	26.8	11.5	1673	20 X77320	Human colon cancer
C 43	26.6	11.4	297	20 Z12750	Human secreted pro
C 44	26.6	11.4	738	20 Z15624	Human gene express
C 45	26.6	11.4	738	20 Z15625	Human gene express

## ALIGNMENTS

RESULT 1

A01024  
ID A01024 standard; cDNA; 278 BP.

XX A01024;

XX A01024;

DT 19-MAY-2000 (first entry)

XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:1015.

XX Human; colon cancer; tumour; diagnosis; gene expression product;  
XX probe; detection; cancerous state; metastasis; identification;  
XX breast cancer; oestrogen receptor-positive breast cancer; therapy;  
XX oestrogen receptor-negative breast cancer; lung cancer; ss.

XX Homo sapiens.

XX WO9958675-A2.

XX 18-NOV-1999.

XX 13-MAY-1999; 99WO-US10602.

XX 14-MAY-1998; 98US-0085426.

XX 15-MAY-1998; 98US-0085537.

XX 15-MAY-1998; 98US-0085696.

XX 21-OCT-1998; 98US-0105234.

XX 27-OCT-1998; 98US-0105877.

XX (CHIR ) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

XX Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

XX Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;





Db 3090 AATTGTAAGTCTGGTAAGCTCTCAACAAAACCTCAGCAATTTCCAGTTTCAT 3038

## RESULT 5

ID Q86896 standard; DNA; 7655 BP.

XX AC Q86896;

XX DT 19-DEC-1995 (first entry)

XX DE S. cerevisiae fks1 gene.

XX KW FK506; antifungal; fungicide; cell wall biosynthesis; fks1 gene; ss.

XX OS Saccharomyces cerevisiae.

XX PN W09510625-A1.

XX PD 20-APR-1995.

XX PF 11-OCT-1994; 94WO-US11498.

XX PR 12-OCT-1993; 93US-0135148.

XX PR 12-OCT-1993; 93US-0135149.

XX PR 12-OCT-1993; 93US-0135150.

XX PA (MERI ) MERCK & CO INC.

XX PI Chrebet GL, Clemas J, Douglas CM, El-Sheerbeini M;

XX PI Foor F, Kahn J, Kelly R, Marrinan JA, Morin NR;

XX PI Onishi JC, Parent SA, Ramadanm, Shei G;

XX DR WPI; 1995-161809/21.

XX DR P-P8DB; R72853.

XX PT DNA encoding FK506 sensitivity genes - used in in vitro assays to

XX PT screen for antifungal cpds., including cpds. that affect cell wall

XX PT biosynthesis

XX PS Disclosure; Fig.6; 130pp; English.

XX SS S. cerevisiae YSK132 (hypersensitive to FK506 at 0.1 ug/ml) was

XX CC crossed to wild-type YFK005 and a meiotic segregant was back-

XX CC crossed to YFK007 (sensitive to 50 ug/ml FK506) to generate strain

XX CC YFK532-10B. The FK506-sensitivity gene FKS1 was then cloned by

XX CC complementation of the FK506 hypersensitive phenotype of

XX CC YFK532-10B. Microbial hosts contg. the DNA are used to identify

XX CC cpds that modulate glucan-synthase activity or that affect

XX CC calcineurin.

XX SQ Sequence 7655 BP; 2188 A; 1440 C; 1511 G; 2516 T; 0 other;

Query Match 12.8%; Score 29.8; DB 16; Length 7655;  
Best Local Similarity 54.0%; Pred. No. 3.3;  
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 13 agtagaataatgttagtggtctcctatgataaggtggaagattgattgaatgaagc 72  
|| ||||| || ||||| || || ||||| || || ||||| || || ||||| || || |||||

Db 4310 AGCGGAATAGATTCTTGGCTCCTCCTCAGTTCAGTCAAGGTGCTCTTCATCCAAAGTAGGC 4251

QY 73 aatctctgattgagacttcagaaacaatgactcggaagctaccgcagat 125  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 4250 AATTGTAAGTCTGGTAAGCTCTCAACAAAACCTCAGCAATTTCCAGTTTCAT 4198

## RESULT 6

ID D00374/c

XX ID D00374 standard; DNA; 5342 BP.

XX AC D00374;

XX XX

DT 09-AUG-2000 (first entry)  
XX DE Rat smooth muscle-alpha actin (SM alphaA) gene regulatory sequence.  
XX KW Smooth muscle alpha actin gene; SM alphaA; rat; smooth muscle cell; SMC;  
KW anticancer; antiproliferative; cardiant; cis regulatory element; cancer;  
KW treatment; proliferative disorder; coronary artery disease; stroke;  
KW hypertension; asthma; multiple gastrointestinal disorder; gene therapy;  
KW urogenital disorder; reproductive disorder; CARG; ds.  
XX OS Rattus sp.  
XX DE Key  
XX FH Location/Qualifiers  
XX FT 1..2558  
XX FT /tag= a  
XX FT 2438..2447  
XX FT /tag= b  
XX FT /label= Carg\_B  
XX FT /note= "Highly conserved cis acting element essential to  
XX FT direct SMC-specific transcription in vivo"  
XX FT 2488..2495  
XX FT /tag= c  
XX FT /label= Carg\_A  
XX FT /note= "Highly conserved cis acting element essential to  
XX FT direct SMC-specific transcription in vivo"  
XX FT 2559..3330  
XX FT /tag= d  
XX FT /number= 1  
XX FT 3331..3656  
XX FT /tag= e  
XX FT /number= 1  
XX FT 3559..3568  
XX FT /tag= f  
XX FT /label= Intronic\_Carg  
XX FT /note= "Highly conserved cis acting element essential to  
XX FT direct SMC-specific transcription in vivo"  
XX WO200024254-A1.  
XX 04-MAY-2000.  
XX 23-OCT-1999; 99WO-US24972.  
XX 23-OCT-1998; 98US-0105330.  
XX (UWVI-) UNIV VIRGINIA PATENT FOUND.  
XX Owens GK, Mack C, Blank R;  
XX WPI; 2000-350519/30.  
XX Isolated polynucleotide is used for inhibiting or treating smooth  
XX muscle cell-related cancer or other proliferative disorders and for  
XX preventing or delaying heart attack  
XX Claim la; Page 84-85; 94pp; English.  
XX The present DNA sequence is the rat smooth muscle alpha actin (SM alphaA)  
XX gene 5' regulatory sequence, comprising the promoter and first intron.  
XX It corresponds to bases -2558 to +2784 of the SM alphaA gene. An intronic  
XX fragment from +773 to +1098 bp relative to the start of transcription is  
XX highly conserved and contains cis-acting elements required to direct  
XX SMC-specific transcription in vivo. This sequence has anticancer,  
XX antiproliferative and cardiant activity. The SM alphaA regulatory  
XX sequence is useful to inhibit and treat smooth muscle cell (SMC)-related  
XX cancers or other proliferative disorders. It is also used for the  
XX treatment of SMC-related disorders, like coronary artery disease, stroke,  
XX hypertension, asthma and multiple gastrointestinal, urogenital and  
XX reproductive disorders. It is also used in gene therapy.  
XX Sequence 5342 BP; 1533 A; 1178 C; 1127 G; 1504 T; 0 other;



Query Match	12.4%	Score 28.8	DB 20	Length 5354
Best Local Similarity	54.8%	Pred. No. 6.1		
Matches 57	Conservative 0	Mismatches 47	Indels 0	Gaps 0
QY 4	tttcagctagtagaataatgttagtgctctctatgatagagtggaaggaattgagttatt 63			
Db 370	TTCATATTTAATAGTATCACTTGTCCACGACAGATATATGCGCCGCTTTTGACACTTGT 311			
QY 64	gaatgaggcaatctctgagttagcttcagacttcagaaaaaatgact 107			
Db 310	GCATTAGCCACAGCAAGACAAAAGGTGTAAAAAACAACGACT 267			
RESULT 8				
ID V74423	/C			
XX V74423	standard; DNA; 18613 BP.			
XX 16-MAR-1999	(first entry)			
XX Staphylococcus aureus	contig SEQ ID #112.			
OS Computer readable medium; vaccine; S.aureus infection; immunodetection;				
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;				
KW skin infection; surgical wound infection; scalded skin syndrome;				
KW toxic shock syndrome; ds.				
XX Staphylococcus aureus.				
OS				
XX Key	Location/Qualifiers			
FT misc_feature	901..960			
FT	/*tag= a			
FT	/note= "these bases represent a line of missing text in			
FT	the sequence listing in the specification. They			
FT	are included to maintain the nucleotide numbering			
FT	given in the specification for this DNA sequence"			
FT misc_feature	2701..2760			
FT	/*tag= b			
FT	/note= "these bases represent a line of missing text in			
FT	the sequence listing in the specification. They			
FT	are included to maintain the nucleotide numbering			
FT	given in the specification for this DNA sequence"			
FT misc_feature	4501..4560			
FT	/*tag= c			
FT	/note= "these bases represent a line of missing text in			
FT	the sequence listing in the specification. They			
FT	are included to maintain the nucleotide numbering			
FT	given in the specification for this DNA sequence"			
FT misc_feature	6301..6360			
FT	/*tag= d			
FT	/note= "these bases represent a line of missing text in			
FT	the sequence listing in the specification. They			
FT	are included to maintain the nucleotide numbering			
FT	given in the specification for this DNA sequence"			
FT misc_feature	8101..8160			
FT	/*tag= e			
FT	/note= "these bases represent a line of missing text in			
FT	the sequence listing in the specification. They			
FT	are included to maintain the nucleotide numbering			
FT	given in the specification for this DNA sequence"			
FT misc_feature	9901..9960			
FT	/*tag= f			
FT	/note= "these bases represent a line of missing text in			
FT	the sequence listing in the specification. They			
FT	are included to maintain the nucleotide numbering			
FT	given in the specification for this DNA sequence"			
FT misc_feature	11701..11760			
FT	/*tag= g			
FT	/note= "these bases represent a line of missing text in			
FT	the sequence listing in the specification. They			
FT	are included to maintain the nucleotide numbering			
FT	given in the specification for this DNA sequence"			





CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 XX  
 SQ Sequence 5333 BP; 1830 A; 921 C; 782 G; 1620 T; 180 other;  
 Query Match 12.2%; Score 28.4; DB 18; Length 5333;  
 Best Local Similarity 50.0%; Pred. No. 8.4;  
 Matches 71; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
 QY 15 tagaataatgttagtgctctatgataggggtggaaggatttgattgattgaatgagga 74  
 DB 3847 TATAGAATTTGGATGGTCTTAACAGAGATGGGTAGACAGCCCTGGATTCTTGGTTA 3788  
 QY 75 tctctgagatgagacttcagaaacaaatgactcgggaagctaccgagattatttt 134  
 DB 3787 TATCGGGTGCAGACGACACACAGCAGCGGGAATACCTTCGTTACAATTTATT 3728  
 QY 135 tggcattgctccatcagttcta 156  
 DB 3727 TGGCATATTGTACATCAATTTTA 3706  
 RESULT 12  
 V31256/c  
 ID V31256 standard; DNA; 17710 BP.  
 XX  
 AC V31256;  
 XX  
 DT 01-OCT-1998 (first entry)  
 XX  
 DE E. coli J96 pathogenicity island contig #70.  
 XX  
 KW PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;  
 KW PAI V; pheV; vaccine; protective immune response; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 PN W09822575-A2.  
 XX  
 PD 28-MAY-1998.  
 XX  
 PF 21-NOV-1997; 97WO-US21347.  
 XX  
 PR 14-OCT-1997; 97US-0061953.  
 PR 22-NOV-1996; 96US-0031626.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (UYWI-) UNIV WISCONSIN.  
 XX  
 PI Choi GH, Dillon PJ, Welch RA;  
 XX WPI; 1998-312461/27.  
 XX  
 PT New isolated uropathogenic E. coli nucleotide sequences - used to  
 PT develop products for the detection of pathogenic E. coli and to  
 PT elicit an immune response to pathogenic E. coli  
 XX  
 PS Claim 21; Page 171-181; 250pp; English.  
 XX  
 CC This sequence represents a E. coli strain J96 contig containing  
 CC pathogenicity island (PAI) sequences, and represents a nucleic acid  
 CC molecule of the invention. PAIs are large fragments of DNA which comprise  
 CC pathogenicity determinants. The sequences of the invention are taken from  
 CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)  
 CC on the E. coli chromosome and is greater than 170 kb. PAI V is located at  
 CC approximately 94 min (at pheR) on the E. coli chromosome and is  
 CC approximately 160 kb in size. Antibodies specific to the proteins encoded  
 CC by the PAI open reading frames of the invention can be used in kits to  
 CC detect uropathogenic E. coli. The proteins are used in vaccines to elicit  
 CC a protective immune response in an animal to the uropathogenic E. coli  
 CC strain J96.  
 XX

SQ Sequence 17710 BP; 4342 A; 4712 C; 4675 G; 3967 T; 14 other;  
 Query Match 12.2%; Score 28.4; DB 19; Length 17710;  
 Best Local Similarity 50.7%; Pred. No. 14;  
 Matches 68; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
 QY 100 caatgactcgggaagctaccgagattattttatttgattgctccatcagctacc 159  
 DB 12783 CGAAGACCAGATAAATCGCGCCCAATTTTGTGATGGTGGTGGATTGCGCGTCC 12724  
 QY 160 atcagatgagagataaagatcatatttgatgagtcctcctcacacacccggttgag 219  
 DB 12723 GACCATCTGACCGCTGGACACATTTTCAGACTCGCGACCACTTCTGCACCGGTGTGAC 12664  
 QY 220 aaggggtattgtg 233  
 DB 12663 GATGGTTCGGTTG 12650  
 RESULT 13  
 V52279/c  
 ID V52279 standard; DNA; 11887 BP.  
 XX  
 AC V52279;  
 XX  
 DT 23-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:146.  
 XX  
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN W09818931-A2.  
 XX  
 PD 07-MAY-1998.  
 XX  
 PF 30-OCT-1997; 97WO-US19588.  
 XX  
 PR 31-OCT-1996; 96US-0029960.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
 PI Kunsch CA, Rosen CA;  
 XX  
 XX WPI; 1998-272225/24.  
 XX  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 XX pneumoniae  
 PS Claim 1; Page 978-985; 1409pp; English.  
 XX  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridise to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and

CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.

SQ Sequence 11887 BP; 3478 A; 2714 C; 2103 G; 3592 T; 0 other;

Query Match 12.1%; Score 28.2; DB 19; Length 11887;  
 Best Local Similarity 49.7%; Pred. No. 13;  
 Matches 72; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
 QY 2 tttttgcagttagtagaataatgttagtgctcctcatgataggggtggaagatttgagttta 61  
 Db 9025 TTTTAAATTCAGTATGGAATGATGGAGATCATATATAAACAATAGAGGTCTTAAGACT 8966  
 QY 62 ttgaatgaggaactctctgagtagagacttcagaaacaatgactcgggaagctaccgc 121  
 Db 8965 TTAGAATAGGCAATAAAGCATATGCAATTCACAAATATCCATGAGCAAAAGCTATAGTCT 8906  
 QY 122 agattattttttggaattgcctc 146  
 Db 8905 AAAATGCTTTCTTTGGCTGTATCTC 8881

RESULT 14  
 V62392  
 ID V62392 standard; DNA; 12565 BP.  
 AC V62392;  
 XX  
 DT 19-JAN-1999 (first entry)  
 DE Human interleukin-1 receptor antagonist gene.  
 XX  
 KW Human; interleukin-1 receptor antagonist gene; IL-1; polymorphism;  
 KW diagnosis; osteoporosis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9844150-A1.  
 XX  
 PD 08-OCT-1998.  
 XX  
 PF 27-MAR-1998; 98WO-GB00944.  
 XX  
 PR 27-MAR-1997; 97GB-0006359.  
 XX  
 PA (GEMI-) GEMINI RES LTD.

XX Keen RW, Spector TD;  
 XX WPI; 1998-557135/47.  
 XX  
 XX Diagnosis of osteoporosis by determining genotype of interleukin-1  
 PT receptor antagonist gene - useful for diagnosing patient  
 PT pre-disposition or susceptibility to osteoporosis and for  
 PT therapeutic intervention  
 XX  
 PS Disclosure; Page 21-27; 36pp; English.  
 XX  
 CC A method has been developed for the diagnosis of osteoporosis comprising  
 CC determining the genotype of an interleukin-1 (IL-1) receptor antagonist  
 CC gene (IL-1RN). The present sequence represents the human interleukin-1  
 CC receptor antagonist gene. The method can be used for the diagnosis of  
 CC disease, including diagnosis of osteoporosis and predisposition or  
 CC susceptibility to osteoporosis and for therapy.

SQ Sequence 12565 BP; 3217 A; 2980 C; 3072 G; 3294 T; 2 other;

Query Match 12.1%; Score 28.2; DB 19; Length 12565;  
 Best Local Similarity 53.1%; Pred. No. 14;  
 Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
 QY 85 tgagacttcagaaaacaatgactcgggaagctaccgcagattattttattgctattgcc 144  
 Db 12129 tgcctctgacattgtagagcttgcacttgagacttgatgaagatgctgtgcc 12188  
 QY 145 tccatcagctaccatcagatgcagatgataaagatcatattgcatgagtc 197  
 Db 12189 tctgctgtctcccaccaggctggagctctgcagagcaggagaaatgactc 12241

RESULT 15  
 X75925  
 ID X75925 standard; DNA; 12565 BP.  
 XX  
 AC X75925;  
 XX  
 DT 29-JUL-1999 (first entry)  
 DE Human interleukin 1RN gene.  
 XX  
 KW Human; interleukin 1; IL-1B; IL-1A; IL-1RN; diagnosis; detection;  
 KW chronic obstructive airway disease; chronic bronchitis; emphysema;  
 KW asthma; chronic bronchiolitis; proinflammatory haplotype; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9924615-A2.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 09-NOV-1998; 98WO-US23721.  
 XX  
 PR 12-JAN-1998; 98US-0005923.  
 PR 07-NOV-1997; 97GB-0023553.  
 XX  
 PA (MEDI-) MEDICAL SCI SYSTEMS INC.

XX Barnes PJ, Duff GW, Giovine M, Lim S;  
 XX WPI; 1999-327420/27.  
 XX  
 XX Genotyping nucleic acid samples for interleukin-1 (IL-1)  
 PT proinflammatory haplotype alleles, useful for predicting  
 PT susceptibility to developing chronic obstructive airway disease  
 XX  
 PS Example 3; Fig 3; 37pp; English.  
 XX  
 CC The present invention describes genotyping a nucleic acid sample from a  
 CC subject to determine at least one allele of an interleukin-1 (IL-1)  
 CC proinflammatory haplotype. A method has also been described for  
 CC determining a subject's susceptibility to developing chronic obstructive  
 CC airway disease (COAD) or for predicting the rapidity or ultimate  
 CC progression of a COAD in the subject by: (a) obtaining a nucleic acid  
 CC sample from the subject; and (b) detecting at least one allele of an  
 CC IL-1 proinflammatory haplotype in the sample, where detection of at  
 CC least one of these alleles indicates that the patient has an increased  
 CC susceptibility to developing COAD. The method is useful for determining  
 CC the susceptibility of subjects to developing chronic obstructive airway  
 CC disease or for predicting the rapidity or ultimate progression of  
 CC chronic obstructive airway disease (COAD). COAD can be asthma, emphysema,  
 CC chronic bronchitis or chronic bronchiolitis. The method provides for  
 CC early identification of chronic obstructive airway disease (COAD),  
 CC facilitating administration of appropriate treatment at the earliest  
 CC stage, thereby increasing the probability of a positive outcome. The  
 CC present sequence represents the human IL-1RN gene.

SQ Sequence 12565 BP; 3217 A; 2980 C; 3072 G; 3294 T; 2 other;

Query Match

12.1%; Score 28.2; DB 20; Length 12565;

Best Local Similarity 53.1%; Pred. No. 14;  
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 85 tgagacttcagaaaaaatgactcggaagctaccgcagagattatttttggcattgcc 144  
Db 12129 tgccttgacattgtagagcttctggcacttgagacttgatgaaagatggctgtgcc 12188  
QY 145 tccatcagttaccatcagatgcgagatgataagatcatttgcattgagtc 197  
Db 12189 tctgcctgtctccccaccaggctgggagctctgcagagcagggaacatgactc 12241

Search completed: November 4, 2000, 13:40:19  
Job time: 16146 sec







RESULT 5  
US-08-603-753D-3  
; Sequence 3, Application US/08603753D  
; Patent No. 5891857  
; GENERAL INFORMATION:  
; APPLICANT: HOLT, JEFFREY T.  
; APPLICANT: JENSEN, ROY A.  
; APPLICANT: PAGE, DAVID L.  
; APPLICANT: KING, MARY-CLAIRE  
; APPLICANT: SZABO, CSILIA I.  
;

```
US-08-603-753D-3
Query Match 11.8%; Score 27.4; DB 3; Length 11283;
Best Local Similarity 65.6%; Pred. No. 8;
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 48 aaggattgattgattgaggaatctctgagtgatgagacttcagaaaaaatgact 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6579 AAGAATTTAAATTAACAATACTTAATGTTGAAGGTGGTTCTTCAGAAAATAATCACT 6638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 108 c 108
Db 6639 c 6639

RESULT 6
US-08-639-501-1
; Sequence 1, Application US/08639501
; Patent No. 5837492
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,501
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/585,391
; FILING DATE: 11-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-8300
; TELEFAX: 202-962-4810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11385 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO

US-09-300-482-1
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 229..10482
US-08-639-501-1

Query Match 11.8%; Score 27.4; DB 3; Length 11385;
Best Local Similarity 65.6%; Pred. No. 8;
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 48 aaggattgattgattgaggaatctctgagtgatgagacttcagaaaaaatgact 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6611 AAGAATTTAAATTAACAATACTTAATGTTGAAGGTGGTTCTTCAGAAAATAATCACT 6670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 108 c 108
Db 6671 c 6671

RESULT 7
US-09-044-946-1
; Sequence 1, Application US/09044946
; Patent No. 6033857
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/044,946
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/639,501
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-8300
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 11385 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 229..10482  
US-09-044-946-1

Query Match 11.8%; Score 27.4; DB 5; Length 11385;  
Best Local Similarity 65.6%; Pred. No. 8;  
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 48 aagdgattgagttatgaatgaggcaatctctgagatgagacttcagaaaaaatgact 107  
DB 6611 AAGAAATTAATTAATCAATTAACACTTAATGTTGAAGTGTCTTCAGAAAAATAATCACT 6670

QY 108 C 108

DB 6671 C 6671

## RESULT 8

US-08-245-294-7/c  
; Sequence 7, Application US/08245294  
; Patent No. 5644047

## GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.  
APPLICANT: Regnery, Russell L.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
TITLE OF INVENTION: DIAGNOSING  
TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA  
TITLE OF INVENTION: INFECTION  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,294  
FILING DATE:

## CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.612  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-9870  
TELEFAX: 404/688-9880

## INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 141..1649

## US-08-245-294-7

Query Match 11.3%; Score 26.4; DB 1; Length 1791;  
Best Local Similarity 59.2%; Pred. No. 7.9;  
Matches 45; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 128 ttatttggcattgcctccatcagtcctaccatcagtcgagatgataagatcatat 187  
DB 1225 TTTGCTAGATCAGGACATCAATTAATCTTCTCACCATTATACCGAAATAATAACATCACCT 1166

QY 188 tgcattgagtcctctt 203

DB 1165 GCCTTGATACCACTT 1150

## RESULT 9

US-08-474-499-7/c  
; Sequence 7, Application US/08474499  
; Patent No. 5693776

## GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.  
APPLICANT: Regnery, Russell L.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE  
TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,499  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/245,294  
FILING DATE: 18-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.612  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-9870  
TELEFAX: 404/688-9880

## INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 141..1649  
US-08-474-499-7

Query Match 11.3%; Score 26.4; DB 2; Length 1791;  
Best Local Similarity 59.2%; Pred. No. 7.9;  
Matches 45; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 128 ttatttggcattgcctccatcagtcctaccatcagtcgagatgataagatcatat 187  
DB 1225 TTTGCTAGATCAGGACATCAATTAATCTTCTCACCATTATACCGAAATAATAACATCACCT 1166



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/755,587  
FILING DATE: 25-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9523959.6  
FILING DATE: 23-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9525555.0  
FILING DATE: 14-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9617961.9  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenneth D Sibley  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5405-135  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1917 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-755-587-1

Query Match 11.3%; Score 26.4; DB 5; Length 1917;  
Best Local Similarity 63.9%; Pred. No. 8.1;  
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 48 aaggattgattgaatgagcaatctctgagtatgagacttcagaaaaaactgaact 107  
Db 176 AAGATTAAATTAATCAATACTTAATGTTGAAGTGTGTTTCAGAAAAATAACT 235  
Qy 108 c 108  
Db 236 C 236

RESULT 13  
US-08-986-485-3/c

Sequence 3, Application US/08986485  
Patent No. 6046030

GENERAL INFORMATION:

APPLICANT: WU, SHUITAN

APPLICANT: SWEET, RAYMOND

APPLICANT: TRONEH, ALEMSEGED

TITLE OF INVENTION: A HUMAN LIG-1 HOMOLOG (HLIG-1)

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: RATNER & PRESTIA

STREET: P.O. BOX 980

CITY: VALLEY FORGE

STATE: PA

COUNTRY: USA

ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/986,485

FILING DATE: 08-DEC-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/059,448

FILING DATE: 22-SEP-1997

ATTORNEY/AGENT INFORMATION:

NAME: PRESTIA, PAUL F

REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70264  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3159 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-986-485-3

Query Match 11.2%; Score 26.2; DB 5; Length 3159;  
Best Local Similarity 49.6%; Pred. No. 12;  
Matches 67; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 83 tatgagacttcagaaaaaactgactcgggaagctaccgcagattatttttggcattg 142  
Db 1776 TATGAAACCCCTGAAACCCCTTATGTACATGATATCAATACATTTTTTGGCTTTT 1717

Qy 143 cctccatcagctaccatcagtatcgagatgataagatcatattgcatgagtcacat 202  
Db 1716 GGATACAAATCCCTCTTGGTTTACTGTGTTTCAGATCCCAAGTCTGTGAACGACT 1657

Qy 203 tcacacaccggttg 217  
Db 1656 GGATACTCCACATGG 1642

RESULT 14

US-08-464-523B-3/c

Sequence 3, Application US/08464523B

Patent No. 5723761

GENERAL INFORMATION:

APPLICANT: Toni A. Voelker

APPLICANT: Ling Yuan

APPLICANT: Jean Kridl

APPLICANT: Deborah Hawkins

APPLICANT: Aubrey Jones

TITLE OF INVENTION: Plant Acyl ACP Thioesterase

TITLE OF INVENTION: Sequences

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Calgene, Inc.

STREET: 1920 Fifth Street

CITY: Davis

STATE: CA

COUNTRY: USA

ZIP: 95616

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 7.1

SOFTWARE: Microsoft Word 5.1(a)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/464,523B

FILING DATE: 02-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/13131

FILING DATE: 10-NOV-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/152,004

FILING DATE: 10-NOV-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/261,695

FILING DATE: 16-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Donna E. Scherer



GenCore version 4.5  
Copyright (c) 1993 - 2000, Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 06:48:03 ; Search time 4352.3 seconds  
(without alignments)  
330.996 Million cell updates/sec

Title: US-09-300-482-1

Perfect score: 233

Sequence: 1 gttttgcagtagtagaat.....ttggacaagggtattgttg 233

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1.\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estom:\*  
68: em\_estov1:\*  
69: em\_estov2:\*  
70: em\_estpl1:\*  
71: em\_estpl2:\*  
72: em\_estpl3:\*  
73: em\_estpl4:\*  
74: em\_estpl5:\*  
75: em\_estro1:\*  
76: em\_estro2:\*  
77: em\_estro3:\*  
78: em\_estro4:\*  
79: em\_estro5:\*  
80: em\_estro6:\*  
81: em\_estro7:\*  
82: em\_estro8:\*  
83: em\_estro9:\*  
84: em\_estro10:\*  
85: em\_estro11:\*  
86: em\_estro12:\*  
87: em\_estro13:\*  
88: gb\_gss1:\*  
89: gb\_gss2:\*  
90: gb\_gss3:\*  
91: gb\_gss4:\*  
92: em\_gss1:\*  
93: em\_gss2:\*  
94: em\_gss3:\*  
95: em\_gss4:\*  
96: gb\_gss5:\*  
97: gb\_gss6:\*  
98: gb\_gss7:\*  
99: gb\_gss8:\*  
100: gb\_gss9:\*  
101: em\_gss5:\*  
102: em\_gss6:\*  
103: em\_gss7:\*  
104: em\_gss8:\*  
105: em\_gss9:\*  
106: em\_gss10:\*  
107: em\_gss11:\*  
108: gb\_gss10:\*  
109: gb\_gss11:\*  
110: em\_gss12:\*  
111: gb\_gss12:\*  
112: gb\_gss13:\*  
113: gb\_gss14:\*  
114: gb\_gss15:\*  
115: gb\_gss16:\*  
116: gb\_gss17:\*

117: gb\_gss18:\*  
 118: gb\_gss19:\*  
 119: em\_gss13:\*  
 120: gb\_gss20:\*  
 121: gb\_gss21:\*  
 122: gb\_gss22:\*  
 123: gb\_gss23:\*  
 124: gb\_gss24:\*  
 125: em\_gss14:\*  
 126: em\_gss15:\*  
 127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164.8	70.7	881	BE216967	BE216967 EST0510 T
2	138	59.2	621	23	AW565200 LG1_328_D
3	131.2	56.3	850	35	BE412598 MCG002.B1
4	99.8	42.8	616	24	AW690515 NF030E09S
5	97	41.6	502	35	BE431846 EST398373
6	97	41.6	560	19	AW031447 EST274901
7	97	41.6	595	20	AW216550 EST295284
8	97	41.6	602	23	BE416585 EST322996
9	97	41.6	608	35	BE449154 EST340253
10	97	41.6	649	11	A191202 EST241911
11	96.6	41.5	287	14	A1941197 eb55g11.Y
12	94.4	40.5	697	23	AW584410 N210452e
13	93.8	40.3	489	24	AW736245 EST332231
14	93.8	40.3	583	19	AV548140 AV548140
15	93.8	40.3	624	19	AV561737 AV561737
16	93	39.9	565	36	BE515829 WHE0606.B
17	92.4	38.7	361	35	BE431906 EST398435
18	92	39.5	538	22	AW455246 EST311906
19	91	39.1	553	25	AW980010 EST310488
20	87.2	37.4	427	19	AV564552 AV564552
21	86	36.9	706	34	BE249133 NF038C11D
22	85.6	36.7	679	24	AW689492 NF019H04S
23	85.2	36.6	372	36	BE523696 M40E8STM
24	79.4	34.1	520	35	BE473149 EST418002
25	78.4	33.6	602	35	BE341181 EST345282
26	77.8	33.4	648	24	AW737079 EST338506
27	76.2	32.7	510	35	BE449284 EST340299
28	76.2	32.7	514	35	BE450103 EST400990
29	76.2	32.7	537	24	AW737104 EST338531
30	70.2	30.1	494	21	AW254983 ML1164 PE
31	69.2	29.7	606	21	AW255222 ML216 DEP
32	67.6	29.0	639	23	AW684970 NF023F11N
33	66.6	28.6	246	36	BE526520 M65J22STM
34	59.6	25.6	479	23	AW617385 EST323796
35	56.4	24.2	515	23	AW617091 EST323502
36	53.8	23.1	657	23	AW685333 NF027G04N
37	52	22.3	382	13	AI894720 EST264163
38	45.4	19.5	436	35	BE342388 EST395232
39	45.4	19.5	486	20	AW223852 EST300663
40	44.8	19.2	496	24	AW774199 EST333350
41	43.8	18.8	588	25	AW980083 EST341594
42	43.6	18.7	412	23	AW621290 EST312088
43	42.4	18.2	316	34	BE203098 EST403120
44	40.2	17.3	539	19	AV526794 AV526794
45	39.8	17.1	629	91	AQ271530 nbxd00261

## ALIGNMENTS

RESULT 1  
 --BE216967

LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BE216967 881 bp mRNA EST 03-JUL-2000  
 EST0510 Triticum aestivum Lambda Zap Triticum aestivum cDNA clone  
 JAL\_5C\_E06\_T3 5' similar to Putative glucose-6-phosphate  
 dehydrogenase, mRNA sequence.  
 BE216967  
 EST.  
 BE216967.1 GI:8904653  
 bread wheat.  
 Triticum aestivum

REFERENCE

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.  
 1 (bases 1 to 881)

AUTHORS

Anderson, J.M., Williams, C.E. and Goodwin, S.B.  
 Analysis of an EST database reveals a probable CF2 resistance gene homolog in wheat

JOURNAL

Unpublished (2000)

COMMENT

Contact: Anderson, J.M.  
 Crop Production & Pest Control Research Unit  
 USDA-ARS

1150 Lilly Hall, West Lafayette, IN 47907, USA

Tel: 765-494-5565

Fax: 765-496-2926

Email: janderson@purdue.edu

Seq primer: T3

High quality sequence stop: 881.

FEATURES

Location/Qualifiers

1..881

/organism="Triticum aestivum"

/strain="P29"

/db\_xref="taxon:4565"

/clone="JAL\_5C\_E06\_T3"

/clone\_lib="Triticum aestivum Lambda Zap"

/tissue\_type="leaf"

/dev\_stage="9 day old seedlings"

254 a 175 c 196 g 256 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;



# TITLE JOURNAL COMMENT

An EST database from Sorghum: light-grown seedlings  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV  
High quality sequence stop: 591  
POLYA-No.

## FEATURES

source

Location/Qualifiers

1..621

/organism="Sorghum bicolor"

/db\_xref="taxon:4558"

/clone\_lib="Light Grown 1 (LGI)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse)

seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI

; The library was made from poly-A RNA in the cloning

vector lambda Zap II. Clones to be sequenced were

prepared by mass excision."

166 a 147 c 133 g 175 t

BASE COUNT  
ORIGIN

Query Match

Best Local Similarity 59.2%; Score 138; DB 23; Length 621;

Matches 181; Conservative 0; Mismatches 45; Indels 6; Gaps 1;

Qy 2 ttttgcagtagtagaataatgttagtgctctatgataggttggaagatttgagttta 61

Db 380 TTTTGGCAATTATATAATATGTCAGTGGTCTCTATGACACTGAGGAAGGATTTCAGAAA 439

Qy 62 ttgaatgaggcaatctctgtagtgagcttcagaaacaaatgactcgggaagctaccgc 121

Db 440 CTGAACAAGACAATATCACAGTATGAGCGCTCA-----ACAAATCAGGAAGCTATCGC 493

Qy 122 agattatttatttggaattgctccatcagtcacccatcagtcagtagatgataaga 181

Db 494 AGGCTCTTTATTGGGATTCCTCCATCTGTCTACCTTCAGTGTGCAAAATGATCAGA 553

Qy 182 tcatttgcagtagtccatcttcacacacgggttgacaaggattttgttg 233

Db 554 ACATATTGCATGATCCATCTTCTCACCCTGGATGACCAAGCTATTTGTTG 605

RESULT 3

BE412598

LOCUS

DEFINITION BE412598 850 bp mRNA EST 24-JUL-2000

CDNA clone MCG002.B10; ITC MCG Barley Leaf/Culm Library Hordeum vulgare

ACCESSION BE412598

VERSION BE412598.1 GI:9410444

KEYWORDS

SOURCE

ORGANISM

Hordeum vulgare

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

1 (bases 1 to 850)

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier

, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,

Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,

Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,

Pecchioni, N., Quiset, C., Schuch, W., Selvaraj, G., Shariflou, M.,

Sorrells, M., Warburton, M. and Wenzel, G.

International Triticaceae EST Cooperative (ITEC): Production of

Expressed Sequence Tags for Species of the Triticaceae

Unpublished (2000)

Contact: Graner A

TITLE

JOURNAL

COMMENT

Institute for Plant Genetics & Crop Plant Research  
Corrensstr. 3, D-06466 Gatersleben GERMANY  
Tel: 49 39482 5521  
Fax: 49 39482 5137  
Email: a\_graner@ipk-gatersleben.de  
International Triticaceae EST Cooperative (ITEC)  
http://wheat.pw.usda.gov/genome.

## FEATURES

source

Location/Qualifiers

1..850

/organism="Hordeum vulgare"

/db\_xref="taxon:4513"

/clone\_lib="MCG002.B10"

/tissue\_type="leaf/culm"

/dev\_stage="etiolated"

BASE COUNT 218 a 214 c 198 g 213 t 7 others

ORIGIN

Query Match

Best Local Similarity 56.3%; Score 131.2; DB 35; Length 850;

Matches 176; Conservative 0; Mismatches 50; Indels 6; Gaps 1;

Qy 2 ttttgcagtagtagaataatgttagtgctctatgataggttggaagatttgagttta 61

Db 443 TTTCTGCAATTATATAATATGTCAGTGGCTCTCTATGACAGTGGAGAAGGTTTGTGAAAAA 502

Qy 62 ttgaatgaggcaatctctgtagtgagcttcagaaacaaatgactcgggaagctaccgc 121

Db 503 CTGAACAAGAAATATCATGATTATGAGATGTCA-----AACACTCAGGAAGCTCCCGT 556

Qy 122 agattatttatttggaattgctccatcagtcacccatcagtcagtagatgataaga 181

Db 557 AGGCTCTTTTACTTGGCATGCTCCATCTGTCTACCTTCAGTGTGCAAAANGATCCGA 616

Qy 182 tcatttgcagtagtccatcttcacacacgggttggaagattttgttg 233

Db 617 ACATATTGCATGAGTCCAACTTCTCGCACTGNATGGACTAGAGTAATTGTTG 668

RESULT 4

AW690515

LOCUS

DEFINITION

AW690515

VERSION

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;

Fabales; Fabaceae; Papilionoideae; Medicago.

1 (bases 1 to 616)

He, X.-Z., Shadle, G., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell

, C.J., Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Dixon

, R.A.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation

Medicago truncatula stem library

Unpublished (2000)

Contact: Dixon RA

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7300

Fax: 580 221 7380

Email: radixon@noble.org

Insert Length: 616 Std Error: 0.00

Plate: 030 row: E column: 09

Seq primer: TCACACAGGAACACGCTATGAC.

Location/Qualifiers

1..616

/organism="Medicago truncatula"





Sat Nov 4 18:11:08 2000

```

Db 394 ACTGGACAAGCAATATCTGAGCAGCAATCTCAAAAAATAGCACAGAAGATCATCCAG 453
Qy 121 cagatttttttgcattgcctccatcagctacccatccatcagtgatgcgagatgataag 180
Db 454 AAGACTTTTCTACTTGTCTCTCCGTCAGTATATCCCTCTCTGTCAGCAATGATAAA 513
Qy 181 atcatattgcagtgatgcctatcttcacacacccggttgacaaagggtttattgttg 233
Db 514 AAGTTATTGTATGAACAAATCTGATCTTGTGTGGTGGACACGAATCGTTGTTG 566

RESULT 9
LOCUS BE449154 608 bp mRNA EST 26-JUL-2000
DEFINITION hirsutum cDNA clone cLHT28E8, mRNA sequence.
ACCESSION BE449154
VERSION BE449154.1 GI:9454657
KEYWORDS EST.
SOURCE Lycopersicon hirsutum.
ORGANISM Lycopersicon hirsutum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 608)
AUTHORS van der Hoeven, R.S., Bezzeredes, J.L., Matern, A.L., Holt, I.E., Liang, F.,
F., Hansen, T., Craven, M.B., Bowman, C.L., Ronning, C.M., Nierman, W.,
Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.
Generation of ESTs from wild tomato (Lycopersicon hirsutum)
trichomes
TITLE Unpublished (2000)
JOURNAL Contact: David Frisch
COMMENT Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source
1..608
/organism="Lycopersicon hirsutum"
/db_xref="taxon:62890"
/clone="cLHT28E8"
/clone_lib="L. hirsutum trichome, Cornell University"
/tissue_type="trichome"
/dev_stage="mixed stages"
/note="Leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yielded a
mixture of cells which is highly enriched for trichome
likely with minor contaminations of other types of leaf
cells"
BASE COUNT 175 a 114 c 131 g 188 t
ORIGIN
Query Match 41.6%; Score 97; DB 35; Length 608;
Best Local Similarity 63.5%; Pred. No. 1.1e-19;
Matches 148; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
Qy 1 gtttttcagttagatgaatatgttagtgctcctatgataggggtggaagatttgattg 60
Db 272 GTTTCGCAACTGATTAATACGTACGTGCTCTTATGATCTCGTGAGGCTTTACGTT 331
Qy 61 attgaatgaggcaatctctcagatgatgagacttcagaaacaatgactcgggaagctaccg 120
Db 332 ACTGGACAAGCAATATCTGAGCAGCAATCTCAAAAAATAGCACAGAAGATCATCCAG 391
Qy 121 cagatttttttgcattgcctccatcagctacccatccatcagtgatgcgagatgataag 180
Db 392 AAGACTTTTCTACTTGTCTCTCCGTCAGTATATCCCTCTCTGTCAGCAATGATAAA 451
Qy 181 atcatattgcagtgatgcctatcttcacacacccggttgacaaagggtttattgttg 233

```

```

Db 452 AAGTTATTGTATGAACAAATCTGATCTTGTGGTGGACACGAATCGTTGTTG 504
Qy 1 gtttttcagttagatgaatatgttagtgctcctatgataggggtggaagatttgattg 60
Db 149 GTTTCGCAACTGATTAATACGTACGTGCTCTTATGATCTCGTGAGGCTTTACGTT 208
Qy 61 attgaatgaggcaatctctcagatgatgagacttcagaaacaatgactcgggaagctaccg 120
Db 209 ACTGGACAAGCAATATCTGAGCAGCAATCTCAAAAAATAGCACAGAAGATCATCCAG 268
Qy 121 cagatttttttgcattgcctccatcagctacccatccatcagtgatgcgagatgataag 180
Db 269 AAGACTTTTCTACTTGTCTCTCCGTCAGTATATCCCTCTCTGTCAGCAATGATAAA 328
Qy 181 atcatattgcagtgatgcctatcttcacacacccggttgacaaagggtttattgttg 233
Db 329 AAGTTATTGTATGAACAAATCTGATCTTGTGGTGGACACGAATCGTTGTTG 381

RESULT 11
A1941197

```

```

FEATURES
source
      1. .287
      Location/Qualifiers
        /organism="Glycine max"
        /db_xref="taxon:3847"
        /clone="GENOME SYSTEMS CLONE ID: Gm-cl010-1821"
        /clone_lib="Gm-cl010"
        /tissue_type="young cotyledons of greenhouse grown plants"
        /dev_stage="2cm long 12 week old"
        /lab_host="Xl10-Gold"
        /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from immature cotyledons (100-200mgs) of old greenhouse
grown plants. The cDNA library was prepared using the
Stratagene pBluescript II SK(+) library construction kit.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly (dT) sequence with a XhoI restriction
site. EcoRI adapters were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA fragments
were directionally cloned into the EcoRI-XhoI restriction
site of the pBluescript vector. The ligated cDNA fragments
were transformed into Xl10-Gold host cells. This library
was constructed by Dr. Lilla Vodkin and Dr. Anu Khanna."
      73 a 43 c 73 g 2 others

```

Query Match	41.5%	Score 96.6;	DB 14;	Length 287;
Best Local Similarity	63.1%;	Pred. No. 1.3e-19;		
Matches 147;	Conservative	0;	Mismatches 86;	Indels 0;
Gaps	0;			
QY	1	gtttttgcagttagagaatattgttagtggtcctctatgattggtggaagatttgagtt	60	
Db	38	gttttttACAGTGTATCAAAATATCTAAAGTGCTCTTATGATTCTGAGGATTCGCTT	97	
QY	61	attgaatgaagcaatctctgagtatgagacttcagaacaaatgactcgggaagtaccg	120	
Db	98	CTTGGNTAAGAGATTTTCAGAGCATGAATATTGTGAAAAGAGTGTGAGGGTTTATCTCG	157	

[illegible]

```

seq primer: 13.
Location/Qualifiers
1. .697
/organism="Medicago truncatula/Glomus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone_lib="MHAM-2K15"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain SOLR"
note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in SOLR cells."
216 a 130 c 146 g 205 t

```

	Query Match	40.5%;	Score 94.4;	DB 23;	Length 697;	
	Best Local Similarity	63.8%;	Pred. No. 7.5e-19;			
	Matches 143; Conservative	0;	Mismatches 81;	Indels 0;	Gaps 0;	
OY	1	gttttgcagttagtagaatattgtagtgctcctcatgatagggtgaagagatttgagtt	60			
Dd	474	GTTTTACAAATGGTGAAATATGTAAGTGGCCCTTATCATTCGGAAGATGGTTTCGCCT	533			
OY	61	attgaatgaggcaatctctctgagtatgagacttcagaaaaacaatgatctcgggaaagctaccq	120			



AV561737	AV561737	624 bp	mRNA	EST	23-JUL-2000
LOCUS	AV561737	Arabidopsis thaliana green siliques Columbia Arabidopsis			
DEFINITION	thaliana	cdna clone SQ156f07F 3', mRNA sequence.			
ACCESSION	AV561737				
VERSION	AV561737.1	GI:8733163			
KEYWORDS	EST				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	1 (bases 1 to 624)				
TITLE	Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.				
JOURNAL	A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 17,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries				
COMMENT	DNA Res. 7, 175-180 (2000) Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: <a href="http://www.kazusa.or.jp/en/plant/">http://www.kazusa.or.jp/en/plant/</a> .				
FEATURES	Location/Qualifiers				
source	1..624				
	/organism="Arabidopsis thaliana"				
	/strain="Columbia"				
	/db_xref="taxon:3702"				
	/clone="SQ156f07F"				
	/clone_lib="Arabidopsis thaliana green siliques Columbia"				
	/tissue_type="green siliques"				
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"				
BASE COUNT	168 a 108 c 153 g 195 t				
ORIGIN					
Query Match	40.3%;	Score 93.8;	DB 19;	Length 624;	
Best Local Similarity	62.7%;	Pred. No. 1.1e-18;			
Matches 146;	Conservative 0;	Mismatches 87;	Indels 0;	Gaps 0;	
QY	1	gttttgcagttagtagaataatgttagtggctctctatgataggtggaaggatttgatt	60		
Db	207	gttttgcagctgatttaagtatgtgagtgcccttattgatctgaggagggttccagag	266		
QY	61	attgaatgaggaactctcgtatgtagacttcagaaacaatgactcgtgggaagtaccg	120		
Db	267	attgacacaggcaatttcagacacgaaatccaaaaatagctactgaagggtctcttag	326		
QY	121	cagattatttatttggccttgcctccatcagctaccaccatcagatcgcagataaag	180		
Db	327	aagactgttttatcttgacacttccaccgctctgtttatcctctgtatgcacgatgcaa	386		
QY	181	atcatattgcagtgccaattctcaacacccggttggacaagggttattgttg	233		
Db	387	gacgtgctgcataacaaatcgtatcttggctgattggacacggattgttgg	439		

Search completed: November 4, 2000, 11:51:59  
Job time: 18236 sec

